



What Makes a Good Trader? On the Role of Intuition and Reflection on Trader Performance

Brice Corgnet, Mark Desantis, David Porter

► To cite this version:

Brice Corgnet, Mark Desantis, David Porter. What Makes a Good Trader? On the Role of Intuition and Reflection on Trader Performance. Working paper GATE 2016-27. 2016. <halshs-01364432>

HAL Id: halshs-01364432

<https://halshs.archives-ouvertes.fr/halshs-01364432>

Submitted on 12 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WP 1627 – September 2016

What Makes a Good Trader?

On the Role of Intuition and Reflection on Trader Performance

Brice Corgnet, Mark DeSantis, David Porter

Abstract:

Using simulations and experiments, we pinpoint two main drivers of trader performance: cognitive reflection and theory of mind. Both dimensions facilitate traders' learning about asset valuation. Cognitive reflection helps traders use market signals to update their beliefs whereas theory of mind offers traders crucial hints on the quality of those signals. We show these skills to be complementary because traders benefit from understanding the quality of market signals only if they are capable of processing them. Cognitive reflection relates to previous Behavioral Finance research as it is the best predictor of a trader's ability to avoid commonly-observed behavioral biases.

Keywords:

Experimental asset markets, behavioral finance, cognitive reflection, theory of mind, financial education

JEL codes:

C92, G2

What Makes a Good Trader?

On the Role of Intuition and Reflection on Trader Performance

Brice Corgnet^{*}
Mark DeSantis^{**}
David Porter^{**}

Abstract

Using simulations and experiments, we pinpoint two main drivers of trader performance: cognitive reflection and theory of mind. Both dimensions facilitate traders' learning about asset valuation. Cognitive reflection helps traders use market signals to update their beliefs whereas theory of mind offers traders crucial hints on the quality of those signals. We show these skills to be complementary because traders benefit from understanding the quality of market signals only if they are capable of processing them. Cognitive reflection relates to previous Behavioral Finance research as it is the best predictor of a trader's ability to avoid commonly-observed behavioral biases.

Keywords: Experimental asset markets, behavioral finance, cognitive reflection, theory of mind, financial education.

JEL CODES: C92, G02.

^{*} EMLYON Business School, Univ Lyon, GATE L-SE UMR 5824, F-69131 Ecully, France

^{**} Argyros School of Business and Economics & Economic Science Institute, Chapman University, Orange, CA, 92866

1. Introduction

The cornerstone models of Finance build upon the assumption that a representative economic agent acts rationally (e.g. Markowitz, 1952; Sharpe, 1964; Samuelson, 1970). The rationality assumption, which requires individuals to correctly apply and carry out mathematical and statistical methods, has been challenged by Behavioral Finance scholars. In contrast to the representative agent model, studies have demonstrated the existence of a large degree of variability in individuals' ability to solve complex problems. When presented with such problems, most individuals tend to rely on simple heuristics instead of performing the requisite calculations (e.g. see Thaler, 1993, 2005; Barberis and Thaler, 2003; Shefrin, 2007 for surveys).

The heterogeneity in individuals' cognitive capacities suggests that we may observe significant differences in their financial decisions. Broader access to the stock market coupled with the increasing complexity of the financial environment (e.g. large number of sophisticated financial instruments, interconnectedness of global markets, etc.) renders the understanding of the relationship between individuals' cognitive capacities and financial decisions both conceptually important and practically relevant.

The goal of our work is to provide an assessment of the skills that predict trader performance in a unified setting. Drivers of trader performance have previously been studied in isolation in different strands of the literature with various types of data (archival, experimental and neuroimaging). Our methodological choice is to use an experimental asset market which allows us to control various aspects of the trading environment (see Bossaerts, 2009; Frydman et al. 2014; Noussair and Tucker, 2014).

Specifically, we choose a market environment that enables us to manage the flow of information into the market so that the effect of individual skills is not confounded with that of insider trading. In our environment, each trader in the market is given a signal regarding the true value of the traded asset (see Plott and Sunder, 1988, henceforth PS).

The experimental methodology also allows us to collect a large set of individual measures for traders. Controlling for an extensive number of individual characteristics (such as IQ, financial

literacy, personality traits, educational background and risk attitudes) ensures the robustness of the effect of our hypothesized predictors.

To derive our main hypotheses, which we test experimentally, we develop a learning model in which the market is populated by traders who differ in terms of cognitive skills. Following the steps of previous works in Behavioral Finance (e.g. Daniel, Hirshleifer and Subrahmanyam, 1998; Rabin, 2002; Eyster, Rabin and Vayanos, 2015) and extending classical market microstructure frameworks (e.g. Glosten and Milgrom, 1985; Easley and O'Hara, 1987), we consider markets in which some traders (fail to) use orders (asset prices, bids and asks) to update their beliefs about asset valuation. We refer to these traders as *reflective (non-reflective)*.

Reflective and non-reflective traders

The literature in cognitive psychology has identified many behavioral biases in beliefs updating (e.g. Kahneman, 2011). A common thread running through each of these biases is the individual's inability to refrain from using automatic responses and simple heuristics which ultimately lead to a failure to apply Bayes' rule (Toplak, West and Stanovich, 2011). Research in cognitive science has identified and validated the cognitive reflection test (henceforth CRT) as a method for capturing a person's ability to avoid common behavioral biases (Oechssler, Roeder and Schmitz, 2009; Toplak, West and Stanovich, 2011, 2014). The CRT consists of questions which all have an appealing and intuitive, yet incorrect, answer (Frederick, 2005). Upon reflection, one may disregard the intuitive answer in favor of the correct one.

Although CRT scores have been found to moderately and positively correlate with general measures of intelligence, only CRT has been found to correlate with an individual's capacity to use Bayes' rule adequately (e.g. Oechssler, Roeder and Schmitz, 2009; Campitelli and Labollita, 2010; Lesage, Navarrete and De Neys, 2013; Toplak, West and Stanovich, 2011, 2014; Sirota, Juanchich and Hagmayer, 2014). Thus, in our setting CRT appears to be the cognitive measure best suited to assess an individual's capacity to update her belief regarding the valuation of the asset. In a market, this capacity is essential to infer other trader's information from orders and thus make more informed and more profitable decisions.

Referring to cognitive reflection as a major dimension of trader performance is also consistent with the fact that CRT-type questions are commonly asked in *Wall Street* interviews for trading positions (Zhou, 2008; Crack, 2014). Not surprisingly, professional traders were found to score remarkably high on the CRT (Thoma et al. 2015).¹

Based on extensive simulations with our learning model, we hypothesize that *reflective* traders who are able to perform Bayesian updating based on orders should earn more than *non-reflective* traders. In addition, our learning model reveals that trader performance not only relies on a trader's capacity to perform Bayesian updating but also on a trader's understanding of the relative proportions of *reflective* and *non-reflective* traders in the market. We refer to those traders who are able to assess with precision the proportion of *reflective* and *non-reflective* traders in the market as *perceptive*.

Perceptive and non-perceptive traders

Perceptive traders better assess the precision of the information contained in orders than *non-perceptive* traders, because they are more likely to know whether orders are placed by *reflective* or *non-reflective* traders. It follows that *perceptive* traders hold more accurate beliefs about the true value of the asset compared to *non-perceptive* traders. We thus expect *perceptive* traders to engage in more profitable trades than *non-perceptive* traders. An important implication of our learning model is that *reflective* and *perceptive* skills are complementary. This holds because traders can only benefit from knowing the precision of market signals if they possess the ability to process such signals.

From an empirical standpoint, we follow previous research on the role of intuition in financial markets (Bruguier, Quartz and Bossaerts, 2010, henceforth BQB) by defining *perceptive* traders as those having the capacity to infer others' intentions (often referred to as theory of mind, e.g. Frith and Frith, 1999). BQB show that trader intuition, measured with the theory of mind eye

¹ In the Experimental Finance literature CRT scores have been found to predict subjects' earnings in experimental asset markets with bubbles (Noussair, Tucker and Xu, 2014; Corgnet et al. 2015).

gaze test (henceforth ToM test, Baron-Cohen et al. 1997), correlates with an individual's ability to predict price changes in experimental asset markets with insiders. They show that individuals with high theory of mind skills are more likely to predict whether prices are set by uninformed traders or insiders. In our markets, all traders start with a signal about the value of the asset so that there are no insiders. However, *reflective* traders are likely to become insiders (or at least more informed than *non-reflective* traders) over the course of the market as they learn other traders' private information from orders. It thus follows from BQB that we should expect high theory of mind traders to more accurately assess whether orders are set by *reflective* or *non-reflective* traders.

In line with our learning model, we find, using experimental asset markets, that both cognitive reflection and theory of mind skills affect trader performance positively. By analyzing traders' orders and shareholdings, we find that traders who are both *reflective* and *perceptive* achieve higher earnings than traders who are *non-reflective* and *non-perceptive* because they are more likely to infer the true value of the asset. We also find evidence for the complementarity of *reflective* (as measured with CRT) and *perceptive* (as measured with the ToM test) skills in line with our model. We show that *perceptive* skills increase earnings the most for *reflective* traders. Finally, we identify situations in which trader skills are especially relevant. For example, we show that *reflective* traders outperform *non-reflective* traders the most when the proportion of *reflective* traders in the market is low. This is most likely due to the fact that there are fewer *reflective* traders competing to seize the wealth of *non-reflective* traders in that case.

2. Contribution to previous literature

This study impacts several related, though not previously connected, strands of the literature, which have considered financial literacy, general intelligence, behavioral biases, theory of mind and self-monitoring as measures of a trader's skills.

Financial literacy

The extensive research on financial education and financial literacy has attempted to identify specific concepts and skills that can enable better financial decisions, especially decisions related

to retirement savings. Financial literacy, which has been defined as the understanding of key financial concepts such as compound interest and the present value of money (Alba and Hutchinson, 1987) has been linked to sound financial decisions such as retirement savings (e.g. Adams and Rau, 2011) and household portfolio diversification (Von Gaudecker, 2015). However, a recent meta-analysis (Fernandes, Lynch and Netemeyer, 2014), shows that the effect of financial literacy is drastically reduced upon the introduction of additional controls, such as personality traits, in the analysis. Our results are consistent with these findings as we show that financial literacy does not significantly affect trader performance when we include relevant control variables in the analysis such as CRT, IQ or personality traits.

General intelligence

Grinblatt, Keloharju and Linnainmaa (2012) also showed, using the same database of Finnish men, that the trades of high-IQ males outperformed those of low-IQ males. These high-IQ individuals exhibited better market timing than their low-IQ counterparts and were more likely to buy stocks whose price was increasing (winning stocks) and sell stocks whose price was decreasing (losing stocks). These findings illustrate that high-IQ traders may have been able to mitigate behavioral biases such as the well-documented disposition effect (e.g. Odean, 1998; Chen et al. 2007). However, the authors leave to further research the study of the exact causes underlying the large returns of high-IQ traders (Grinblatt, Keloharju and Linnainmaa, 2012). Our study extends these works by assessing the effect of IQ in a controlled environment in which we can pinpoint the sources of high trader performance. We show that IQ, measured using Raven scores, is likely to have a positive and significant effect on trader performance because it significantly correlates with our two main drivers of trader performance: cognitive reflection and theory of mind.

Behavioral finance, heuristics and biases

The Behavioral Finance literature has grown by demonstrating how the heuristics and biases identified by cognitive psychologists (e.g. Tversky and Kahneman, 1974) can distort financial decisions. Behavioral biases that have been shown to correlate with trading behavior include

overconfidence as well as an individual's failure to understand random sampling. Our work contributes to this literature by demonstrating that cognitive reflection, which has been shown to be the best predictor of people's ability to avoid behavioral biases (e.g Oechssler, Roider and Schmitz, 2009; Toplak, West and Stanovich, 2011, 2014), significantly affects trader earnings. This finding adds to the Behavioral Finance literature by directly connecting trader performance to a specific cognitive skill.

Theory of mind and trader intuition

The work of BQB highlights the role of trader intuition as an important determinant of trader performance. BQB were able to identify a brain area (paracingulate cortex) that activates when insider trading is present in the market. The authors link this brain area to trader intuition and, more specifically, to an individual's theory of mind skills. Their study showed that theory of mind skills as measured by, for example, the eye gaze test correlate with one's ability to predict price changes in experimental asset markets with insiders. The authors did not, however, study the relationship between trading behavior and theory of mind skills as they chose to focus on forecasting abilities:

“Of course, thinking about prices and forecasting them are integral to successful trading, but these two steps alone leave out the actual placing of orders. Trading intuition concerns not only assessment of what is going on in the market and prediction of future prices, but also submission of the right orders. Our study only considers the first two facets; future work should shed light on the third.”

Bruguier, Quartz and Bossaerts (2010), p. 1721

Our work seeks to address this third facet of successful trading by assessing the relationship between a trader's performance and theory of mind skills.

Self-monitoring and miscalibration

Our analysis also builds upon the work of Biais et al. (2005) as we used a similar trading environment (see PS) to study the effect of individual characteristics on trader performance. The authors focus on self-monitoring (a measure of one's disposition to attend social cues, Snyder

and Gangestad, 1986) as well as overconfidence which was measured using a *miscalibration* test (e.g. Russo and Schoemaker, 1992). They conjectured that individuals who attain high scores on the self-monitoring scale would behave more strategically than others, as they would more accurately infer other traders' signals from asset prices. By contrast, overconfidence would inhibit traders' learning of the true valuation of the asset. Ultimately, the authors posited that traders who perform well on both the self-monitoring scale and *miscalibration* scales should earn higher trading profits in market situations in which the inference regarding the true value of the asset is particularly challenging. They identify these market situations as winner's curse traps in which traders may be induced to hold an incorrect belief of the true value of the asset. They confirmed these conjectures in an oral double auction trading experiment. Our work supports Bias et al. (2005) by showing that certain market situations such as winner's curse traps are key to understanding trader performance. Our results also confirm that a crucial dimension of trader performance is one's capacity to adequately infer information from orders. Using both a learning model and laboratory experiments, we extend Bias et al. (2005) by identifying two independent but complementary drivers (operationalized using CRT and ToM tests) of a trader's capacity to update beliefs correctly.

3. Design

3.1. Asset markets

Our experimental asset market environment is similar to the experimental design of Plott and Sunder (1988, henceforth PS). The only notable difference with their original design is that our study uses a computerized instead of an oral continuous double auction. We made this choice purposefully as computerized continuous double auction trading mechanisms are widely used in actual stock market exchanges (Parsons et al. 2008) and in online prediction markets (www.predictit.org).² The PS design is suited to our study as it implements a rich environment including both public and private information. A trader's ability to update her belief of the asset

² In continuous double auctions, traders can submit, at any time, offers to buy or sell the asset. Traders can also accept other traders' existing orders to buy or sell an asset.

value by using her private information in conjunction with the information inferred from other traders' orders is the key element of our study of the determinants of individual trader performance.

We use the same parameters as in the PS study (Market 9, Treatment C) regarding possible asset values, endowments and the number of market replications. Each experimental session consisted of 17 market periods during which participants could trade an experimental asset whose exact value was not known with certainty. This asset could assume one of three possible values: 50, 240 or 490 francs. A franc was worth \$0.001. At the beginning of each market period, every trader was informed of a possible value the asset could not take. As half of the traders were given one hint (e.g. "Not 50") and the other half were given the other possible hint (e.g. "Not 240"), the aggregate information available to all traders in the market was complete. PS argue that in a rational expectation equilibrium prices should reflect the true value of the asset (e.g. 490). In this paper, we focus on individual trading behavior rather than on the aggregation of information at the market level.³ More specifically, we focus on the relationship between trader performance and individual characteristics. We assess performance by studying traders' earnings as well as their orders and shareholdings.

3.2. Protocol

We recruited a total of 204 participants from a subject pool of more than 1,500 students at a major Western US University. We conducted a total of 17 sessions, each with 12 traders. In the ten baseline sessions, traders were endowed with 1,200 francs in cash and 4 shares. To ensure our results are not artifacts of this specific endowment structure, two sessions were run using the PS approach of endowing each subject with a 25,000 franc loan that had to be repaid at the end of each period and 4 shares. We added a third treatment to assess the robustness of our findings

³ We leave the study of information aggregation at the market level to Corgnet, DeSantis and Porter (2015a), which includes a detailed analysis of this topic.

to higher stakes by doubling both the traders' initial cash endowment (to 2,400 francs) as well as the value of shares (to 100, 480 and 980).⁴

Before each session started subjects completed a 10-minute training quiz regarding the probability of occurrence of each possible asset value (50, 240 or 490 francs) at the end of each of the 17 markets. This training (see online Appendix O1, Instructions Part 1) consisted of having subjects predict the outcome of a spinning wheel over 10 trials. Each correct prediction was rewarded 25 cents, and each incorrect answer incurred a 10 cent penalty. Average earnings (including a \$7 show-up fee) for the two and one-half hour experiments were equal to \$48.0 in our ten baseline sessions, \$33.5 in our two loan sessions and \$86.3 in our five high stakes sessions. We summarize our experimental design in Table 1.

Table 1: Summary of our experimental design

<i>Treatment</i>	<i>Number of traders</i>	<i>Number of markets (market length in minutes) -Sessions-</i>	<i>Endowment / Loan Francs (Assets)</i>	<i>Asset values Francs (Probabilities)</i>	<i>Trading mechanism</i>
Baseline	12	17 (5) - 10 -	1,200 (4)	50, 240, 490 (0.35,0.45,0.20)	Computerized continuous double auction
Loan	Same	17 (5) - 2 -	25,000 Loan (4)	Same	Same
High Stakes	Same	17 (5) - 5 -	2,400 (4)	100, 480, 980 (0.35,0.45,0.20)	Same

3.3. Surveys

As part of the experiment, participants answered questionnaires regarding various psychological and cognitive traits on two separate occasions. At the conclusion of the last market period, subjects participated in a brief end-of-experiment survey. Then, one month after the original study they were invited to complete a more comprehensive survey.

⁴ In this treatment, we also made changes in the writing of the instructions (see online Appendix O1 for the instructions and Appendix F5 for a separate analysis of the high stakes treatment).

One-hour follow-up lab survey

All of the subjects who participated in our experiments were invited one month after the original study to complete a one-hour survey as part of the laboratory policy to collect individual information about subjects who are registered in the pool. A large majority of participants completed the one-hour survey (82%, 167 out of 204).⁵ Importantly, the subjects who attended the survey do not significantly differ from the subjects who did not in terms of trader earnings (p -value = 0.16, Wilcoxon rank-sum test, WRS henceforth).

The survey was computerized, and, as is common practice in the literature, the tests were not incentivized though subjects earned a \$15 flat fee. The survey captured our two main measures of trader performance, cognitive reflection and theory of mind, through the CRT and ToM test.

Cognitive reflection test

We administered the extended (seven-question) version of the CRT in which the original three questions (Frederick, 2005) are augmented with four additional questions recently developed and validated by Toplak, West and Stanovich (2014) (see Appendix A for details). Our measure of cognitive reflection is given by the total number of correct answers (from 0 to 7). The Cronbach alpha reliability score for the extended CRT (0.70) is in line with that of Toplak, West and Stanovich (2014) who reported a reliability of 0.72. Subjects had 5 minutes to complete the CRT.

Theory of mind test

Following BQB and De Martino et al. (2013), we administered the ToM test (Baron-Cohen et al. 1997) to assess subjects' theory of mind skills. In this task, participants looked at images of people's eyes and had to choose one of four feelings that best described the mental state of the person whose eyes were shown. Our ToM score is defined as the number of correct answers to the 36 question, 10-minute test. In Appendix A, we present the remaining variables included in the follow-up survey.

⁵ Out of the 167 subjects who completed the survey, 15 completed the survey more than two months after the completion of the market experiments.

End-of-experiment survey

In addition, at the end of each session, subjects answered demographic questions and completed a series of cognitive tests. The duration of the survey was 25 minutes. These tasks were computerized and the tests were not incentivized though subjects earned a \$3 flat fee for completing the survey. The survey included tests for financial literacy and self-monitoring. Participants also took the CRT and the Raven test exactly as in the follow-up survey.⁶

In the five high stakes sessions, we also included a Bayesian updating test taken from Charness and Rabin (2009). In four of the five high stakes sessions, we also asked subjects to assess their session rank on the CRT (see Appendix D2 for the corresponding analysis). We describe in detail the tests used in the end-of-experiment survey in Appendix A.

4. Model

To derive our main hypotheses, we develop a learning model that closely mimics our experimental market. One important aspect of the model is the inclusion of a heterogeneous population of agents. The model relates to previous market microstructure research by partially relying on Bayesian learning to describe how traders infer information from orders. We extend previous microstructure research by studying markets in which some traders may not use Bayes' rule or may not correctly assess the proportion of traders using Bayes' rule. Our model extends both market microstructure and behavioral finance models by (1) considering the case in which traders have disperse information (see PS) and (2) allowing them to learn not only from asset prices but also from buy and sell orders. Our work closely relates to market microstructure models in which traders' strategies directly follow from their beliefs about the value of the asset (e.g. Glosten and Milgrom, 1985; Easley and O'Hara, 1987). Traders buy (sell) the asset whenever a sell offer (a buy offer) is below (above) their current belief of the true value of the asset. In the

⁶ One advantage of measuring CRT twice is that it allowed us to conduct our CRT analysis for the whole sample instead of considering the subsample of 82% of participants who came back for the follow-up survey (see e.g. Table 3). Another advantage of measuring CRT and Raven twice is that it enables us to show that subjects who did not participate in the follow-up survey did not significantly differ from those who did participate with respect to these two cognitive tests (p -values > 0.20 , WRS). Due to time constraints, we did not add the theory of mind test to the end-of-experiment survey.

case in which all traders learn the true value of the asset, our model reaches a situation in which prices equal the rational expectation equilibrium prediction. In that case, traders would only trade the asset at its true value similar to a fully-revealing rational expectation equilibrium in which traders' beliefs coincide with the equilibrium price of the asset. In our model, asset prices do not instantaneously converge to the true value of the asset. Even in the most favorable case in which the market is populated solely by traders who are *reflective* (and all traders know that all other traders are *reflective*), it takes on average three orders for traders' beliefs to converge to the true value of the asset and for prices to trade at this exact value.⁷ Given the complexity of the model in which traders infer the true value of the asset using all orders, we use simulations to derive our main hypotheses (see Section 4.4).⁸

4.1. Trading and information

We follow our experimental design in modeling the trading process. In particular, we assume prices occur as a result of a continuous flow of bids and asks posted by traders. Our model is based on a decentralized auction mechanism in which all traders (1) have the same portfolio of cash and shares and (2) receive a hint about the true value of the asset.⁹

At the beginning of each market, a trader (selected at random) posts a bid-ask spread. A second trader is then selected and given the option to either accept the current bid or ask or improve the bid-ask spread.¹⁰ Traders (selected at random) continue to improve the bid-ask spread until a trade occurs.¹¹ This sequential trading in which one unit of the asset is traded at a time resembles the

⁷ This is calculated using the simulations presented in Section 4.4. In our model, learning cannot be immediate because traders update their beliefs only after observing an order.

⁸ Relatedly, Pouget (2007) used simulations to study learning in markets with insiders. Our approach differs, however, because traders learn about the true value of the asset by inferring other traders' information through orders. In Pouget (2007) traders learn to play certain strategies.

⁹ A market maker (or a specialist) is commonly utilized in the market microstructure literature (e.g. Glosten and Milgrom, 1985; Easley and O'Hara, 1987). In order to model our experimental design as closely as possible we do not include a market maker in our model.

¹⁰ A bid-ask spread is only updated when the newly selected trader has enough cash (shares) to cover the bid (ask) position. In line with our experimental markets, traders are not permitted to sell short or borrow funds.

¹¹ The sequence of trades is not random, however. Traders either will or will not trade at a given bid-ask spread depending upon their current belief regarding the true value of the asset.

classical Glosten and Milgrom (1985) model. If a trade occurs, then the price, p , is set at the current best bid (or ask) and the trading book is updated.¹²

The decision of traders to improve the bid-ask spread or trade depends on their beliefs regarding the true asset value. Specifically, traders will improve the current bid-ask spread if their beliefs lie within the current spread. In that case, the updated bid (ask) will be drawn from a uniform distribution between the current bid (ask) and the trader's belief. If traders' beliefs are below (above) the current best bid (ask), then they would sell (buy) the asset. Trading thus follows, similar to our experimental design, a bid-ask improvement rule.

In our market environment, traders observe market prices as well as bids and asks. We will thus assume that traders can use all orders as information to update their beliefs. The way they update their beliefs depends on their behavioral type.

4.2. Behavioral types

We define behavioral types based on traders' learning limitations. The first limitation follows from a failure to perform Bayesian updating (e.g. Charness and Levin, 2009) (see Section 4.2.1 below) and the second follows from a failure to gauge other traders' behavioral types (e.g. Bazerman and Samuleson, 1983) (see Section 4.2.2 below).

4.2.1. Reflective and non-reflective traders

In line with quasi-Bayesian learning models proposed by Daniel, Hirshleifer and Subrahmanyam (1998) and Rabin (2002), our framework considers people who differ in their ability to apply Bayes' rule to new information.¹³ Unlike these prior studies, we consider a model in which

¹² The accepted bid/ask is removed from the book, and the next best bid/ask, if one exists, becomes the current best bid/ask.

¹³ Daniel, Hirshleifer and Subrahmanyam (1998) accounts for stock market under- and over- reaction to news by developing a model in which traders overvalue the precision of their private information and update their beliefs self-servingly by downplaying information which may not be consistent with their prior beliefs. Rabin (2002) considers the case in which investors assign too much weight to small samples to account for stock market under- and over- reaction to news. Both Daniel, Hirshleifer and Subrahmanyam (1998) and Rabin (2002) build their models on extensive cognitive psychology literature showing that people largely fail to apply Bayes' rule suffering instead from a number of common and long-lasting biases (e.g. Tversky and Kahneman, 1974; Stanovich, 2009b; Khaneman, 2011). Payzan-LeNestour and Bossaerts (2015) provides a decisive argument challenging Bayesian

heterogeneous agents interact in the market. In particular, we assume our traders to either be *reflective* or *non-reflective*. *Reflective* traders apply Bayes' rule to infer the true value of the asset as they observe orders (asset prices, bids and asks). *Non-reflective* traders only update their beliefs once, on the basis of their private information, and do not use orders to infer other traders' private information. These identifying characteristics are consistent with two important models in the literature: the prior information model (Lintner, 1969) and the cursed equilibrium model (Eyster and Rabin, 2010; Eyster, Rabin and Vayanos, 2015). According to the former, traders make decisions based solely on their private information and fail to infer other traders' information from market prices. Interestingly, Corgnet, DeSantis and Porter (2015a) report extensive support for this model using a wealth of experimental asset market data. The behavior of traders in the prior information model closely relates to the widely-documented phenomenon of the winner's curse (Bazerman and Samuelson, 1983; Thaler, 1988, 1991). The winner's curse occurs when auction participants fail to anticipate the informational content of other peoples' bids and end up paying too much for the auctioned item. Applied to financial markets, Eyster, Rabin and Vayanos (2015) refer to *cursed* traders as those who "neglect the informational content of prices." Our model extends their work by providing a psychological profile, which crucially hinges on cognitive reflection, for *cursed* traders.

As the CRT is a key determinant of an individual's capacity to properly use Bayes' rule, it is an appropriate measure of a trader's ability to infer other traders' information from prices. In line with previous works, we report that CRT scores correlate positively and significantly with Bayesian updating scores ($r = 0.30$, $p\text{-value} = 0.02$, $n = 60$).^{14,15}

4.2.2. Perceptive and non-perceptive traders

learning models by showing that people may not be naturally inclined to learn like Bayesians even when getting extensive opportunities to do so, such as in armed bandit tasks.

¹⁴ We also confirm that our general measure of intelligence, the Raven test, correlates with Bayesian updating scores less strongly than CRT scores do ($r = 0.22$, $p\text{-value} = 0.09$, $n = 60$). In addition, when regressing Bayesian updating scores on both CRT and Raven scores, we report that only the coefficient associated to CRT scores reaches statistical significance ($p\text{-value} = 0.01$ for CRT scores, compared to $p\text{-value} = 0.12$ for Raven scores) (see Appendix D1).

¹⁵ Along with CRT, we also use a Bayesian updating performance measure which was originally proposed by Charness and Levin (2009) (see Appendix A).

In our market environment, traders may incorrectly update their beliefs after observing orders either because they are not able to apply Bayes' rule or because they hold wrong beliefs about other traders' behavioral types. For example, *reflective* traders who believe that other traders are *non-reflective* will downplay the value of orders as accurate signals of other individuals' information compared to traders who believe others are *reflective*.

For the sake of our model, we define *perceptive* (*non-perceptive*) traders as those who know (do not know) the actual proportion of *reflective* traders in the market on trader performance. Following previous research on the role of intuition in financial markets (BQB), we view *perceptive* traders as those who have high theory of mind skills and are thus more likely to know others' intentions and behavioral types. The role of theory of mind skills in capturing one's ability to assess the level of sophistication of others also relates to a series of works linking theory of mind with an individual's ability to reduce strategic uncertainty in two-player games (e.g. McCabe et al. 2001; Hampton et al. 2008; Ekins et al. 2013; Kimbrough, Robalino and Robson, 2016). These works stress that theory of mind skills are crucial in understanding and anticipating other players' choices in games. In our setup, traders with high theory of mind skills will be more likely to infer others' private information from their orders as they have a better assessment of their behavioral types than traders with low theory of mind skills.

Non-reflective traders, whether they are *perceptive* or not, do not use orders to update their beliefs regarding the valuation of the asset. Thus, we consider three distinct behavioral types: *reflective* and *perceptive*, *reflective* but *non-perceptive*, and *non-reflective*.

Importantly, the measures we use for defining *reflective* (CRT) and *perceptive* (ToM) traders are not correlated ($r = -0.03$, $p\text{-value} = 0.704$). This is in line with our model which assumes that *reflective* and *perceptive* are two independent behavioral types. It follows that a *reflective* trader is equally likely to be either *perceptive* or *non-perceptive*.

4.3. Learning

In our model, learning occurs after each market event. Market events either correspond to an improvement of the bid-ask spread or to a transaction. Based on these market events, *reflective*

traders update their beliefs of the true value of the asset by applying Bayes' rule to infer other traders' information. By contrast, *non-reflective* traders do not update their beliefs relying entirely on their private information.¹⁶

4.3.1. Prior beliefs and hints

In line with our experimental market environment, traders receive a hint $h \in \{\text{"Not 50"}, \text{"Not 490"}, \text{"Not 240"}\}$ before the market starts. We denote by $\mu_{n,h}^\tau$ the belief of a trader of behavioral type τ after the n^{th} market event and given hint h , where $\tau \in \{R_P, R_{NP}, NR\}$ and R_P (R_{NP}) [NR] stands for a *perceptive reflective* (*non-perceptive reflective*) [*non-reflective*] trader. We will refer to all *reflective* traders, whether they are *perceptive* or *non-perceptive*, as R .

It follows, applying Bayes' rule that prior beliefs (before any market event takes place, that is $n = 0$) of all trader types are as follows: $\mu_{0, \text{Not 50}}^\tau = 316.9$, $\mu_{0, \text{Not 240}}^\tau = 210$ and $\mu_{0, \text{Not 490}}^\tau = 156.9$ for any $\tau \in \{R_P, R_{NP}, NR\}$.

4.3.2. Traders' orders and beliefs updating

We illustrate how traders update prior beliefs based on market events with an example. We consider a market solely populated by *reflective* traders in which the value of the asset is 240 (see Appendix B for a more general description of the updating procedure and for a more detailed example). We consider three successive market events.

First market event: bid-ask spread {200, 360}

The first market event always corresponds to a trader setting a bid-ask spread. In our example, we consider this first bid-ask spread to be $\{b_1 = 200, a_1 = 360\}$, where b_n (a_n) denote the bid (ask) associated to the n^{th} market event. This bid-ask spread leads all *reflective* traders to update

¹⁶ Our framework can be extended by considering the general case in which *non-reflective* traders do not necessarily ignore orders, which is an extreme version of *cursed* trading (Eyster, Rabin and Vayanos, 2015). For example, one might define traders to be *reflective*-type η if they need to observe η orders consistent with a certain value of the asset before updating their belief accordingly. This extension of our model would relate to the works stressing the prominent role of inattention in financial decisions (e.g. Agnew, Balduzzi, and Sunden, 2003; Andersen et al. 2015). This also relates to Payzan-LeNestour and Bossaerts (2015) who suggest that individuals who have the ability to apply Bayes' rule do not because of their lack of vigilance.

their beliefs upwards (see Figure 1, first market event). This is the case because a bid of 200 could not have been posted by traders who hold the hint “Not 490” because they value the asset at 156.9. This leads traders to increase their belief that the value of the asset could be 490 thus increasing their belief regarding the valuation of the asset. At this point, *reflective* traders who are *non-perceptive* update their beliefs similarly to those who are *perceptive* because both *reflective* and *non-reflective* traders hold the same initial prior beliefs given a hint. *Non-reflective* traders do not update beliefs based on market events.

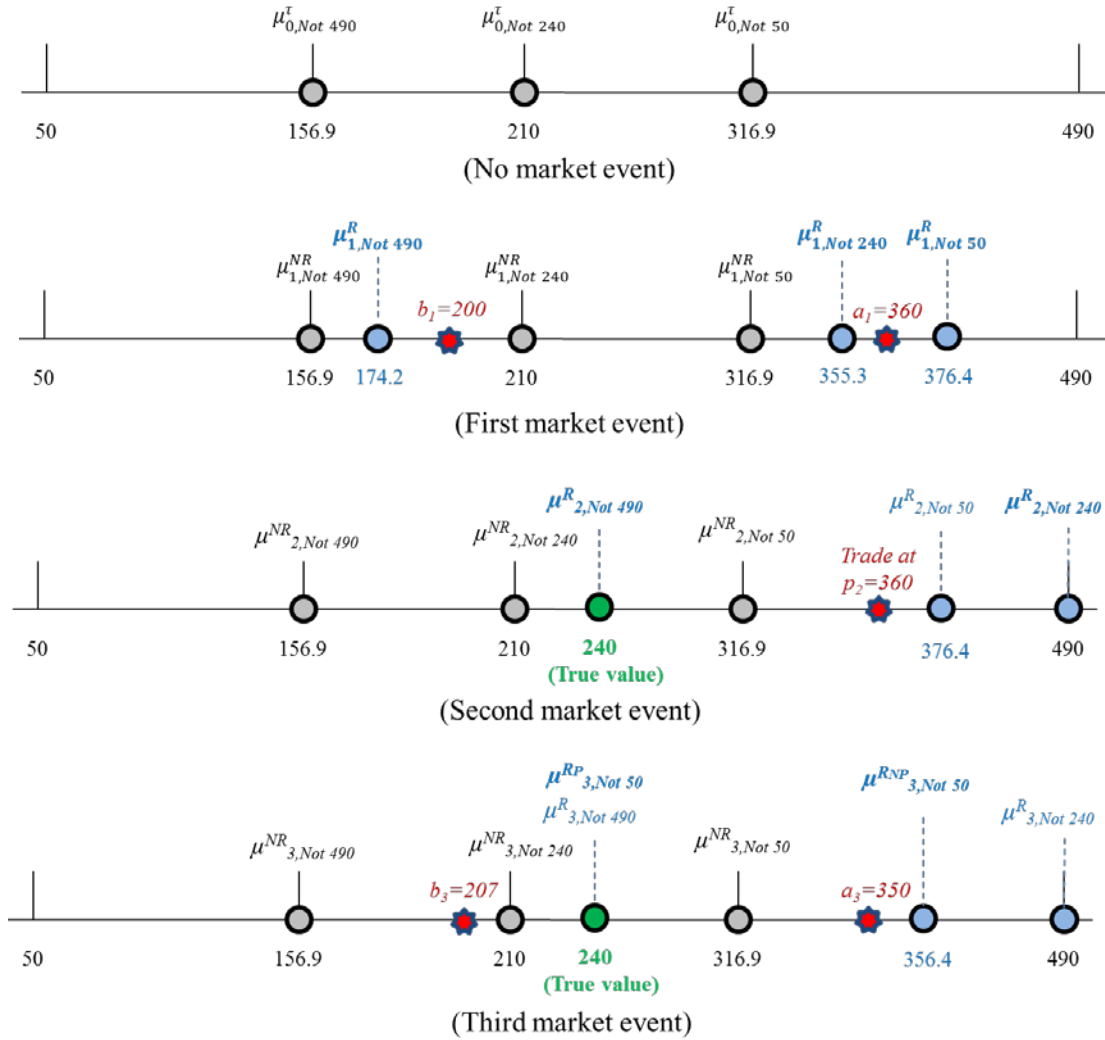


Figure 1. Orders and evolution of traders' beliefs (example).

Second market event: trade at 360

Suppose the second market event corresponds to a transaction at the current ask price ($p_2=360$, where p_n denotes the price at which the asset is traded in the n^{th} market event). Given traders' beliefs, only *reflective* traders who hold the hint "Not 50" would have been buying the asset at 360 because they are the only type of trader who value the asset at more than 360 ($\mu_{1,Not\ 50}^R = 376.4$) (see Figure 1, first market event). It thus follows that all *reflective* traders learn from this second market event that the value of the asset cannot be 50. *Reflective* traders who hold the hint "Not 490" thus infer that the value of the asset must be 240 ($\mu_{2,Not\ 490}^R = 240$). However, *reflective* traders who already hold the hint "Not 50" learn nothing from this second market event ($\mu_{2,Not\ 50}^R = \mu_{1,Not\ 50}^R$). Given that the true value of the asset is 240, traders who hold the hint "Not 240" do not exist. However, *reflective* traders must keep track of the evolution of these traders' beliefs because they do not know the true value of the asset at the beginning of the market and thus are unaware that these traders do not exist. Traders who hold the hint "Not 240" would infer from the second market event that the value of the asset is 490.

Third market event: bid-ask spread {207, 350}

Suppose the third market event consists of a trader posting a bid-ask spread $\{b_3 = 207, a_3 = 350\}$ which improves the current bid of 200 (which has not yet been taken) and provides a new ask (the previous ask of 360 has already been accepted). The only traders that can update beliefs at this point are the *reflective* traders who hold the hint "Not 50" because they have not yet discovered the true value of the asset. These *reflective* traders who are also *perceptive* know that the market is populated solely with *reflective* traders. Thus, they know that an ask of 350 is not compatible with the value of the asset being 490. This is the case because, had the value of the asset been 490, they would know that the traders populating the market (*reflective* traders holding either the hint "Not 50" or "Not 240") value the asset at more than 350 ($\mu_{2,Not\ 50}^R = 376.4$ and $\mu_{2,Not\ 240}^R = 490$) (see Figure 1, second market event). It follows that *perceptive reflective* traders holding the hint "Not 50" learn that the true value of the asset is 240 (see Figure 1, third market event). However, *reflective* traders who hold the hint "Not 50" do not infer the

true value of the asset from the posted ask of 350 if they are not *perceptive*. This is due to the fact that *non-perceptive* traders wrongly believe that some *non-reflective* traders populate the market in which case an ask of 350 is still compatible with the value of the asset being 490. This is so because *non-reflective* traders holding the hints “Not 50” and “Not 240” value the asset at less than 350. For example, if *non-perceptive* traders believe the proportion of *reflective* traders in the market is 25% (4 out of 12), then their belief regarding the true value of the asset is 356.4 after observing the third market event.

This example illustrates the learning procedure by which *reflective* traders can uncover the true value of the asset from orders and it shows why *non-perceptive reflective* traders may learn more slowly than *perceptive reflective* traders.

4.4. Simulations and hypotheses

We conduct simulations of the learning model to derive a series of hypotheses that can be tested experimentally. Each simulation is conducted with the same number of traders as in our experimental asset markets (12) and the same endowment of cash (1200 francs) and shares (4).¹⁷

4.4.1. Simulation procedures

Simulations were run for various combinations of three exogenous model parameters. The first parameter corresponds to the value of the asset $\in \{50, 240, 490\}$. The second parameter is the proportion of *reflective* traders ($\alpha \in \{\frac{2}{12}, \frac{3}{12}, \dots, 1\}$) in the market.¹⁸ This proportion defines the ratio of each behavioral type in the market because we split *reflective* traders equally between the *perceptive* and *non-perceptive* types.¹⁹ The third parameter is the *non-perceptive reflective* traders’ belief regarding the proportion of *reflective* traders in the market ($\alpha^{RNP} \in \{\frac{0}{12}, \frac{1}{12}, \dots, 1\}$) where $\alpha^{RNP} \neq \alpha$). For each of the 429 combinations of our three parameters (v, α, α^{RNP}) we perform

¹⁷ Our results are, however, robust to considering smaller or bigger markets as well as different endowments. These simulations are available upon request from the authors.

¹⁸ We need at least $\alpha = \frac{2}{12}$ so that all three behavioral types are represented.

¹⁹ For an odd number of *reflective* traders, each simulation randomly determines whether there is one more *perceptive reflective* trader or *non-perceptive reflective* trader in the market. Our results are robust to considering more uneven distributions of *perceptive* and *non-perceptive* traders.

25,000 simulations. Each simulation runs until 5 trades have been executed. Our results are, however, robust to different stopping rules.²⁰

We derive three hypotheses assessing differences in earnings, orders and shareholdings across behavioral types.

4.4.2. Hypotheses

In our model, traders who obtain the highest earnings are those who are able to most accurately learn others' information from orders. As illustrated in Section 4.3, learning hinges on one's capacity to be both *reflective* and *perceptive*. In addition, *perceptive* skills are only useful to those traders who attempt to uncover others' information from orders. This implies that *perceptive* skills will only increase the earnings of *reflective* traders. That is, *perceptive* and *reflective* skills are complementary.

In addition, the model points to situations in which trader skills are likely to be most relevant. This is the case, for example, when the market is populated by a low proportion of *reflective* traders. In that case, the competition between *reflective* traders for seizing *non-reflective* traders' wealth is milder leading to higher earnings for *reflective* traders. Another important situation in which trader skills are especially relevant is the case in which traders receive an imprecise signal regarding the true value of the asset. This occurs when traders receive the hint "Not 240" in which case the signal does not provide clear guidance regarding the asset value (50 or 490). By contrast, the hints "Not 50" and "Not 490" provide clearer guidance to traders regarding the true value of the asset. In statistical terms, the signal "Not 240" is less precise than the signals "Not 50" and "Not 490".²¹ In our model, the less precise a trader's hint is the more there is to learn from orders about other traders' hints. It follows that trader skills should matter the most when traders' private information consists of imprecise signals of the true value of the asset ("Not 240") than more precise signals ("Not 50" or "Not 490").

²⁰ For example, our results are robust to stopping the simulations after 30 trades which is the average number of trades observed in our experimental asset markets (results available from the authors).

²¹ Given our market parameters (see Table 1), a trader receiving the hint "Not 240" believes the asset value can be either 50 or 490 with probability 63.6% and 36.4%. The standard deviation of the "Not 240" trader posterior belief is thus equal to 211.7 compared to 115.4 [94.3] in the case of "Not 50" ["Not 490"] traders.

Based upon our model's predictions, we make the following hypothesis regarding traders' earnings.

Hypothesis 1 (Earnings).

- i) *Reflective traders will earn more than non-reflective traders. This difference will be larger when the market is populated by a larger proportion of non-reflective traders.*
- ii) *Reflective traders who are perceptive will earn more than those reflective traders who are not perceptive. The difference in earnings between perceptive reflective and non-reflective traders will be larger for traders who receive an imprecise hint ("Not 240") than for those traders who receive a more precise hint ("Not 50" or "Not 490").*
- iii) *Reflective and perceptive skills will be complementary in explaining trader earnings.*

Table B3.1 (Panel A) in Appendix B3 reports the results of our simulations regarding earnings comparisons across behavioral types supporting Hypothesis 1. In Figure B3.1 in Appendix B3, we also motivate the second part of Hypothesis 1i.

In our model, high earners are those traders whose beliefs are more in line with the true value of the asset. We define an order to be *consistent* with the true value of the asset when a bid or a purchase (an ask or a sale) is lower (higher) than or equal to the true value of the asset. Hypothesis 2 is structured to identify the impact of *reflective* and *perceptive* skills on the proportion of *consistent* orders submitted to the market. Similar to Hypothesis 1, we also expect the difference in the proportion of *consistent* orders between *perceptive reflective* traders and *non-reflective* traders to be most pronounced when they receive imprecise signals.²² We thus utilize our model to make the following hypothesis regarding traders' orders *consistency*.

Hypothesis 2 (Orders).

- i) *Reflective traders will place a greater proportion of orders (bids, asks, purchases and sales) that are consistent with the true value of the asset than non-reflective traders.*

²² We do not make predictions regarding market composition because markets populated by a large proportion of *non-reflective* traders may impede the learning process and thus lead to lower order consistency than markets populated by a small proportion of *non-reflective* traders.

- ii) *Reflective traders who are perceptive will place a greater proportion of orders that are consistent with the true value of the asset than those reflective traders who are not perceptive. The difference in the proportion of consistent orders between perceptive reflective and non-reflective traders will be larger for traders who receive an imprecise hint (“Not 240”) than for those traders who receive a more precise hint (“Not 50” or “Not 490”).*
- iii) *Reflective and perceptive skills will be complementary in explaining the proportion of consistent orders.*

Table B3.1 (Panel B) reports the results of our simulations regarding the *consistency* of traders’ orders across behavioral types and hints. These results support Hypothesis 2.

In our model, high earners are more likely to learn the true value of the asset and thus hold more (less) shares at the end of the market when the asset is undervalued (overvalued). In particular, we expect skilled traders (those who are both *reflective* and *perceptive*) to hold less (more) shares than unskilled traders when the value of the asset is lower (higher) than the market price, which is the case when the value of the asset is 50 (490).²³

As above, our model enables us to make the following hypothesis regarding shareholdings.

Hypothesis 3 (Shareholdings).

- i) *Reflective traders will hold less [more] shares at the end of the market than non-reflective traders when the actual value of the asset is low (50) [high (490)].*
- ii) *Reflective traders who are perceptive will hold less [more] shares at the end of the market than those reflective traders who are not perceptive when the actual value of the asset is low (50) [high (490)].*
- iii) *Reflective and perceptive skills will be complementary in explaining traders’ shareholdings.*

²³ In our model, observed prices can never be below 50 or above 490.

Table B3.2 reports the results of our simulations for shareholdings across behavioral types and hints. These results support Hypothesis 3.

5. Experimental results

We analyze our experimental data for evidence supporting Hypotheses 1-3. To that end, our empirical definition of a *reflective (non-reflective)* trader is one who scores in the top (bottom) quartile on the CRT. We define *perceptive (non-perceptive)* traders as those who score in the top (bottom) quartile on the ToM test.

Online Appendix O2 includes figures displaying the evolution of asset prices across the 17 markets for each of the 17 sessions we conducted.

5.1. Statistical method

To investigate our hypotheses, we use linear panel regressions with random effects in which each subject is treated as a cross-section observation ($n = 204$) and each market as a time observation ($t = 17$).^{24,25} In each regression, we report robust standard errors clustered at the session level. We add treatment dummy variables controlling either for the type of endowment which was given to subjects (Loan Dummy) or for the stakes (High stakes Dummy). In the regressions using trader earnings as the dependent variable, the coefficients associated to the Loan Dummy should be negative because subjects earned less on average in those sessions (\$33.5) than in the baseline sessions (\$48.0) in which they were given a \$1.2 endowment each period. For the high stakes sessions, we divide subjects' earnings by two so that average earnings are in line with those in the baseline. High stakes sessions closely resemble our baseline sessions when considering aggregate market variables (see Figure F5.1 in Appendix F5). For example, the mean absolute deviation of average prices with respect to the true value of the asset (the main measure of information aggregation used in PS) does not significantly differ between the

²⁴ Fixed effects cannot be used since we are assessing the effect of time-invariant regressors (such as CRT) on trader performance. Also, using the Breusch-Pagan Lagrange Multiplier test, we reject the hypothesis that random effects are not appropriate (p-values < 0.001 across all specifications used in the results section).

²⁵ In lieu of analyzing trader wealth by market, we could use average earnings across markets as our dependent variable. Despite substantially reducing the number of available observations, the use of average earnings yields results that are remarkably similar to those obtained via our panel data analysis (see Appendix F3).

baseline and high stakes sessions (p-value = 0.27, WRS). In Appendix F5, we also report no significant effect of high stakes sessions on our individual analysis of trader performance. For the sake of the main analysis, we thus pool the high stakes sessions with the other sessions.

We also control for the actual value of the asset in a given market (Asset value) as well as for the hint received by the subject in a given market (“Not 50”, “Not 240” or “Not 490”). We expect regression coefficients associated to Asset value to be positive because subjects earned more in markets in which the value of the asset was higher. We control for gender by using a dummy variable that takes value one if the trader is a male and value zero otherwise (Male Dummy). We finally control for subjects’ IQ scores (Raven, 1941) and financial literacy scores (Fernandes, Lynch and Netemeyer, 2014). All individual measures are standardized to facilitate direct comparisons of the magnitudes of the coefficients associated to each of these variables.

Given the moderate but significant correlation between CRT, ToM, Raven and financial literacy scores (see Table 2 below), we use orthogonalization techniques to deal with collinearity issues (Golub and Van Loan, 1996). The pattern of correlation in Table 2 is consistent with previous research.²⁶

Table 2. Correlation matrix for individual cognitive measures.

	CRT	ToM	Raven
ToM	-0.030	–	–
Raven	0.298****	0.211***	–
Financial literacy	0.353****	0.076	0.192***

*p-value<0.10, **p-value<0.05, ***p-value<0.01 and ****p-value<0.001

n =204 for all pairwise correlations except those involving ToM in which case *n* = 167.

We orthogonalize Raven and financial literacy scores with respect to CRT scores in all regressions in which ToM scores are not included. Likewise, we orthogonalize CRT, Raven and financial literacy scores with respect to ToM scores in all regressions in which both CRT and ToM scores are included.

²⁶ Indeed, the significant positive correlation between Raven and CRT is consistent with Frederick (2005) whereas the significant positive correlation between Raven, CRT and financial literacy is also found in the financial literacy research (e.g. Fernandes, Lynch and Netemeyer, 2014). Finally, the significant positive correlation between Raven and ToM scores is in line with Ibanez et al. (2013).

In Appendix F, we provide extensive robustness checks, none of which are found to affect the nature of our findings. In Appendix F1, we add further individual controls to our regression analyses such as risk attitudes, educational background (school attended and grades) and personality traits (big five personality traits and self-monitoring scale). Session fixed effects are considered in Appendix F2. In Appendix F3, we define trader performance measures for the entire experiment abstracting away from the panel structure of our dataset. We also conduct additional panel regression analyses controlling for the possible serial correlation in the error term. In Appendix F4, we assess the robustness of our findings to the case in which dependent variables are not orthogonalized to deal with collinearity issues. The robustness of our findings to only using high stakes data is confirmed in Appendix F5. In Appendix F6, we conduct our analyses by defining the main predictors as deviations from the session average.

5.2. Earnings

In line with Hypotheses 1i and 1ii, we find that both CRT and ToM scores related positively to trader earnings. In Figure 2, we compare the earnings of traders in the bottom and top quartiles in terms of CRT and ToM scores.²⁷ We observe that both CRT and ToM scores have a positive effect on earnings. The earnings gap is the largest when comparing traders in the bottom quartile (\$34.9) of both CRT and ToM scores with traders in the top quartile (\$40.1). A trader who is in the top quartile of both CRT and ToM earns more than the top 25% of earners (\$39.8) whereas a trader who is in the bottom quartile of both CRT and ToM earn as little as the bottom 25% earners (\$34.5). We can also show that traders who are in the top 25% in both CRT and ToM outperform those who are in the top 25% according to either CRT (p-value = 0.045, WRS) or ToM scores (p-value = 0.046, WRS).

²⁷ These quartiles are determined using participants scores in the current study. The quartile definitions are identical for both CRT (scores above 4 for top 25% and below 2 for bottom 25%) and ToM (scores above 28 for top 25% and below 25 for bottom 25%) if we use the whole pool of subjects registered at the lab instead of the participants in the current study.

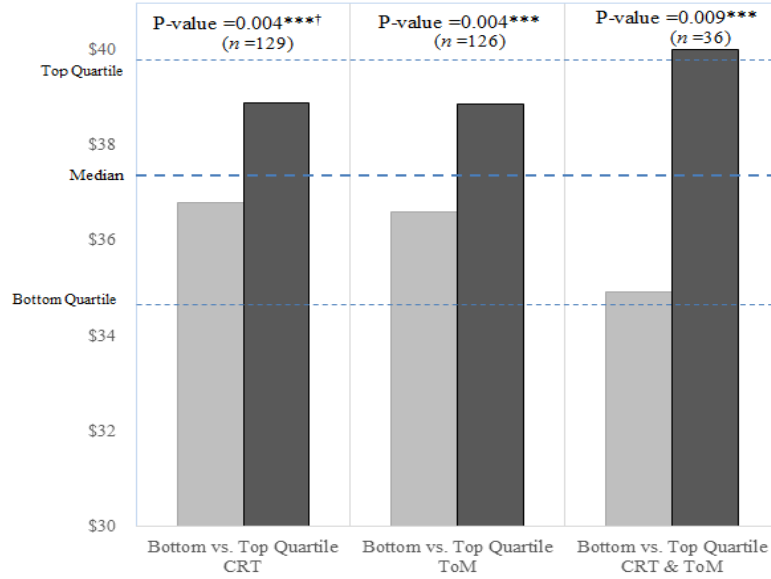


Figure 2. Average trader earnings for participants in the bottom and top quartiles for CRT and ToM scores. We also represent the bottom quartile, median and top quartile for the whole distribution of participants' earnings (see dashed horizontal lines).

†P-values based on WRS comparing earnings of traders in the bottom and top quartiles of CRT, ToM and both CRT and ToM. *p-value<0.10, **p-value<0.05, ***p-value<0.01 and ****p-value<0.001

Hypothesis 1i (Reflective traders)

Using panel regressions that control for market variables as well as specific characteristics of the individual traders, we show that, in line with Hypothesis 1i, CRT scores positively affect earnings (see CRT score in column [1] of Table 3). *Reflective* traders outperform *non-reflective* traders in terms of CRT (see CRT Top 25% Dummy in column [2], which takes value one if a trader's CRT score is in the top 25% of participants).

To test the second part of Hypothesis 1i, we calculate for each of the 17 sessions the ratio of bottom quartile CRT traders to top and bottom quartile CRT traders. We divide the sample between sessions which have a higher ratio of *non-reflective* traders than the median session and

those which have a lower ratio than the median session.²⁸ We show that top quartile CRT traders outperform those in the bottom quartile only when the ratio of *non-reflective* traders in the market is higher than the median session (see column [3] and [4]) in Table 3).

Table 3. Trader earnings as a function of individual characteristics and market variables

<i>Earnings (in \$)</i>				
Sample:	All	Top & Bottom 25% CRT	High ratio of <i>non-reflective</i> traders	Low ratio of <i>non-reflective</i> traders
	[1]	[2]	[3]	[4]
Intercept	1,076.524**** (47.652)	1,010.852**** (62.468)	981.552**** (85.544)	1,021.871**** (67.547)
<i>Hypothesized predictors</i>				
CRT score	44.850*** (16.272)	—	—	—
CRT Top 25% Dummy	—	85.423** (43.272)	117.146** (57.684)	86.519 (68.122)
<i>Controls</i>				
<i>Individual characteristics</i>				
Raven score	35.281 (22.599)	34.694 (29.476)	30.791 (36.241)	25.777 (21.229)
Financial literacy score	4.505 (12.771)	2.922 (22.136)	45.569* (25.112)	36.793 (28.230)
Male Dummy	37.665* (33.173)	84.157** (40.756)	142.097** (64.116)	41.159 (46.891)
<i>Market characteristics</i>				
Loan Dummy	-879.488**** (23.809)	-829.89**** (59.955)	-718.661**** (47.747)	-869.719**** (15.527)
Hint “Not 50” Dummy ²⁹	191.144**** (48.822)	196.661**** (52.959)	227.945*** (76.905)	153.945** (70.625)
Hint “Not 490” Dummy	211.045**** (20.222)	189.80**** (28.288)	202.978**** (44.845)	171.137**** (31.191)

²⁸ Given that we conducted an uneven number of sessions (17), we cannot exactly split the two samples in half. We report results for the case in which there is one more session classified as having a high ratio of *non-reflective* traders.

²⁹ Both hint dummy variables (“Not 50” and “Not 490”) affect earnings positively and significantly in line with the fact that they offer a more precise signal than the hint “Not 240”.

Asset value	3.834**** (0.132)	3.908**** (0.149)	3.871**** (0.251)	3.951**** (0.132)
Observations	n = 3,468	n = 2,193	n = 1,258	n = 935
Prob > χ^2	0.000	0.000	0.000	0.000
R ²	0.645	0.650	0.663	0.642

*p-value<0.10, **p-value<0.05, ***p-value<0.01 and ****p-value<0.001

In line with our model, markets which are populated by a larger ratio of *non-reflective* traders tend to exhibit higher levels of mispricing, defined as the absolute difference between the average market price and the true value of the asset. Indeed, mispricing is significantly higher in markets that have a higher ratio of *non-reflective* traders than the median market than in markets that have a lower ratio than the median market (137.26 vs. 125.17, p-value < 0.001, WRS). Thus, showing that *reflective* traders perform particularly well in sessions in which the ratio of *non-reflective* traders is high implies that *reflective* traders tend to do well in markets in which mispricing is high. We purposefully state Hypothesis 1i in terms of the proportion of *non-reflective* traders instead of the level of mispricing in the market because this proportion is exogenously determined in our model whereas mispricing is endogenous.

Mispricing is also high in specific situations identified by Biais et al. (2005) as winner's curse traps. Following Biais et al. (2005), we define winner's curse traps as situations in which the true value of the asset is not 240, and the participant's signal does not rule out 240. Winner's curse traps manifest themselves in that context when prices are close to 240. In that case, traders are likely to wrongly infer that the true value of the asset is 240. The level of mispricing in Winner's curse traps situations is significantly larger than otherwise (202.12 vs. 132.39, p-value < 0.001, WRS). In Appendix C1, we show that, in line with the Biais et al. (2005) findings, the positive effect of trader's skills (CRT and ToM scores) tends to be most pronounced in Winner's curse traps situations.

Hypothesis 1ii and 1iii (Reflective and perceptive traders)

In Table 4, we provide regression results that test Hypotheses 1ii and 1iii. We report a significantly positive effect of ToM scores on earnings in line with Hypothesis 1ii (columns [1]

to [5]). In particular, *perceptive* traders outperform *non-perceptive* traders (see ToM Top 25% Dummy in column [3]).

Consistent with Hypothesis 1ii, we show that theory of minds skills positively affect the earnings of *reflective* traders (see ToM scores coefficient in column [4]) whereas this effect is much less pronounced for *non-reflective* traders (see ToM scores coefficient in column [5]).

Table 4. Trader earnings as a function of individual characteristics (including ToM) and market variables

<i>Earnings (in \$)</i>					
Sample:	All	All	Top & Bottom 25% ToM	Top 25% CRT	Bottom 25% CRT
	[1]	[2]	[3]	[4]	[5]
Intercept	1,054.421**** (45.403)	1,053.817**** (47.519)	1,061.810**** (94.285)	957.624**** (82.885)	974.442**** (134.504)
<i>Hypothesized predictors</i>					
ToM score	39.082*** (13.464)	41.560**** (12.847)	—	77.033**** (23.195)	46.116** (19.489)
ToM Top 25% Dummy	—	—	128.402**** (31.568)	—	—
ToM score × CRT score	—	13.555 (12.802)	—	—	—
CRT score	44.517** (17.595)	44.616*** (16.790)	43.633** (21.366)	142.996** (63.166)	1.457 (112.975)
<i>Controls</i>					
<i>Individual characteristics</i>					
Raven score	45.040** (18.698)	46.822** (18.749)	40.319 (27.110)	41.871 (30.279)	44.211** (23.270)
Financial literacy score	8.398 (14.511)	8.605 (14.535)	5.432 (20.400)	-5.117 (23.095)	35.593 (26.075)
Male Dummy	74.916** (35.923)	76.286** (36.374)	115.988** (46.132)	24.505 (56.777)	147.881*** (52.012)
<i>Market characteristics</i>					
Loan Dummy	-865.255**** (32.963)	-864.649**** (32.365)	-817.738**** (31.041)	-971.786**** (42.289)	-663.644**** (73.364)
Hint “Not 50” Dummy	198.245**** (45.520)	198.100**** (45.556)	211.801*** (66.598)	137.5868* (75.777)	222.133**** (59.192)
Hint “Not 490” Dummy	234.134**** (16.264)	234.198**** (16.360)	225.702**** (40.319)	169.027**** (34.530)	218.423**** (46.904)
Asset value	3.808****	3.808****	3.485****	4.000****	3.765****

	(0.135)	(0.135)	(0.252)	(0.180)	(0.243)
Observations	n = 2,839	n = 2,839	n = 1,513	n = 1,020	n = 816
Prob > χ^2	0.000	0.000	0.000	0.000	0.000
R ²	0.650	0.651	0.644	0.667	0.661

*p-value<0.10, **p-value<0.05, ***p-value<0.01 and ****p-value<0.001

In line with Hypothesis 1ii, we also show that being a skilled trader (i.e., being *reflective* and *perceptive*) is most consequential when traders receive an imprecise hint (“Not 240”) (see Appendix C2). In a regression using trader earnings as the dependent variable, we find that the coefficient for the *perceptive reflective* trader dummy variable (which takes value one when a trader is both *perceptive* and *reflective*) reaches statistical significance (p-value < 0.01) when traders receive an imprecise hint (see Table C2.1, column [1] in Appendix C2). By contrast, the *perceptive reflective* trader dummy fails to reach standard statistical significance when traders receive more precise hints (p-value = 0.089) (see Table C2.1, column [2] in Appendix C2).

Finally, we report directional evidence for Hypothesis 1iii as the interaction effect between ToM and CRT scores is positive (see ToM score \times CRT score coefficient in Table 4 column [2]), but not statistically significant (p-value = 0.290). However, this ToM \times CRT interaction effect is positive and significant when considering traders’ orders as well as shareholdings.

5.3. Orders

In line with the previous earnings results, we show that both CRT scores and ToM scores explain the *consistency* of traders’ orders. We illustrate our findings graphically below.

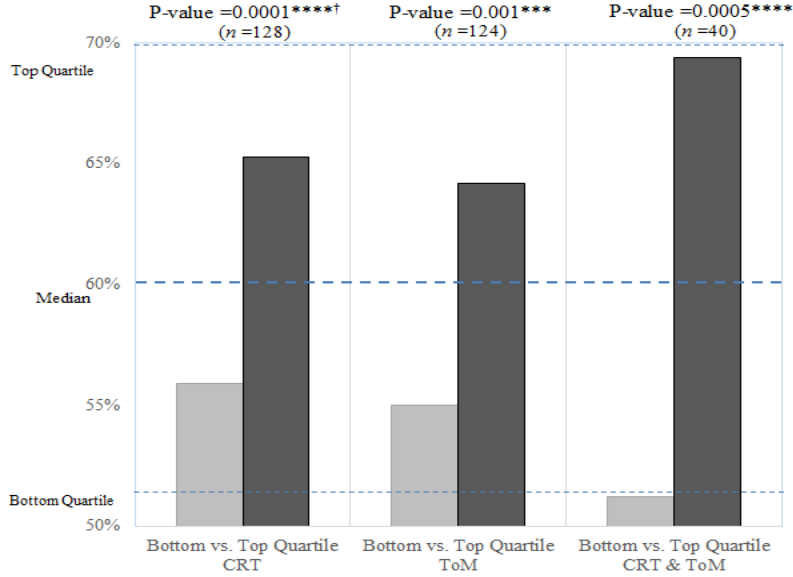


Figure 3. Average proportions of *consistent* orders for participants in the bottom and top quartiles for CRT and ToM scores. We also represent the bottom quartile, median and top quartile for the whole distribution of the proportion of *consistent* orders (see dashed horizontal lines).

†P-values based on WRS comparing the proportion of *consistent* orders of bottom and top quartiles traders in terms of CRT, ToM, and both CRT and ToM. *p-value<0.10, **p-value<0.05, ***p-value<0.01 and ****p-value<0.001

In line with Hypothesis 2i, we find that *reflective* traders place a greater proportion of *consistent* trades than *non-reflective* traders (see Table C3.1 in Appendix C3). Confirming Hypothesis 2ii, we also report that *perceptive reflective* traders place more *consistent* orders than *reflective* traders who are not *perceptive* (see column [4] in Table C3.2 in Appendix C3). In line with our model, we find that the positive effect of ToM scores is limited to *reflective* traders (see columns [4] and [5] in Table C3.2). Importantly, we find, in line with Hypothesis 2iii, a positive and significant interaction effect between CRT and ToM scores in explaining the proportion of *consistent* trades (see column [2] in Table C3.2). Also in line with Hypothesis 2ii, we find that *perceptive reflective* traders place a significantly greater proportion of *consistent* orders than *non-reflective* traders when considering traders who receive the imprecise hint “Not 240” (p-value = 0.013, see Appendix C2, Table C2.1, column [3]). This is not the case when considering

only traders who receive the more precise hints (“Not 50” or “Not 490”) (p-value = 0.254, see Appendix C2, Table C2.1, column [4]).

We further assess the consistency of traders’ behavior with the true value of the asset by studying final shareholdings.

5.4. Shareholdings

In Figure C4.1 in Appendix C4 we compare shareholdings across behavioral types when the asset value is either 50 or 490. In line with Hypothesis 3i, *reflective* traders tend to hold less (more) shares at the end of the market than *non-reflective* traders when the asset value is low (high). The differences are not statistically significant, however. Hypothesis 3ii is supported as *perceptive reflective* traders hold less (more) shares than *non-perceptive reflective* traders when the asset value is low (high). This difference is only statistically significant when the true value of the asset is high. The lack of statistical significance of some of these basic comparisons may follow from the fact that we are considering separately low and high asset values leading to lesser statistical power. Thus, to highlight differences, we study shareholdings across behavioral types using the data from all market periods and combining the observations from markets in which the value of the asset is either 50 or 490. To that end, we define our dependent variable as being the net purchases [net sales] of certificates of traders in markets in which the value of the asset is 490 [50].³⁰ This dependent variable captures the extent to which shareholdings at the end of the market are *consistent* with the true value of the asset. We refer to this variable as net holdings consistency. Following Hypotheses 3i and 3ii, we expect *perceptive reflective* traders to buy shares when the true value of the asset is high (490) and sell shares when the true value of the asset is low (50). We thus expect these traders to be characterized by a high level of net holdings consistency.

In Table C4.1 (column [1]) in Appendix C4, we show that net holdings consistency is positively affected by CRT scores in line with Hypothesis 3i. In line with Hypothesis 3ii, we

³⁰ Net purchases [sales] of shares are defined as: shareholdings at the end of the market - 4 [4 - shareholdings at the end of the market] where 4 reflects the initial endowment of shares.

show a significantly positive effect of ToM scores on net holdings consistency for *reflective* traders (see column [3]) which disappears for *non-reflective* traders (see column [4]). In addition, we observe, in line with Hypothesis 3iii, a positive and significant interaction effect between CRT and ToM scores in explaining net holdings consistency (see Table C4.1, column [2] in Appendix C4).

In Appendix D, we provide further analyses that test the soundness our model hypotheses. We also provide a separate analysis of gender effects in Appendix E.

6. Conclusion: On the Role of Behavioral and Experimental Finance in Financial Education

Our work shows that cognitive reflection and theory of mind skills are the main drivers of trader performance. Commonly-studied variables such as financial literacy, personality traits and risk attitudes play a lesser role in understanding trader performance. Our findings echo recent research in cognitive psychology stressing the limitations of current cognitive tests (e.g. Raven) to assess rational thinking, which is defined as one's capacity to avoid behavioral biases and thus apply Bayes' rule correctly (Stanovich, 2009). The author advocates the development of a rationality quotient (RQ) that would complement the commonly-measured IQ (Stanovich, 2016). To date, the best measure of RQ is provided by the CRT (Toplak, West and Stanovich, 2014). Our finding regarding the predominant role of RQ as opposed to IQ (Raven test) in explaining trader performance is consistent with the well-known observation of Warren Buffet (Loomis, 2013, p. 101):

“You don't need a rocket scientist. Investing is not a game where the 160 IQ guy beats the guy with the 130 IQ... Rationality is essential...”

Our results also support the recent move toward introducing Behavioral Finance findings in the standard Finance curriculum. For example, even though the CFA (Chartered Financial Analyst) curriculum emphasizes financial literacy, numeracy and computational ability, it has recently included a Behavioral Finance section.³¹ Given the many limitations and challenges people face

³¹ See <http://www.businessinsider.com/cfa-level-3-behavioral-finance-2013-5>

in undoing their own behavioral biases (e.g. Larrick, 2004), an important research topic is to gauge whether people can learn to avoid these biases the same way they learn financial concepts such as compounding and the time value of money.

An interesting avenue for future research would be to design financial training programs that emphasize the importance of using orders as valuable signals about other traders' private information. Such programs would, in addition to alerting participants to the negative financial consequences of behavioral biases, teach them how to avoid these biases.

Our findings also show that cognitive reflection is not sufficient to achieve maximal trader performance thus leaving room for financial training based on social cognitive skills such as theory of mind. This would thus pave the way for a novel financial training program based on social cognitive skills.

In that respect, we believe the experimental methodology could play a fundamental role by engaging people in market environments that reproduce relevant aspects of real markets without endangering their own wealth. In addition, another issue with learning financial principles in real markets is the large number of factors affecting asset prices that blur the individual participant's responsibility in poor performance. Clear and unambiguous feedback on one's own trading performance is a very appealing argument for the use of controlled market environments as a tool for financial education.³² Experience in these experimental markets could help subjects learn to suppress their behavioral biases and properly update their beliefs based on orders. Subjects might also learn to properly assess the proportion of robots belonging to each of the three behavioral types we have identified.

7. References

Adams, G. and B. Rau (2011) "Putting off Tomorrow to do What you Want Today: Planning For Retirement," *American Psychologist* 66(3), 180-192.

See http://www.cfainstitute.org/utility/pages/search_results.aspx?k=behavioral%20finance&s=All%20Sites

³² The computer software used for these experiments is open source and available for free. It can be run on virtually any browser and any electronic device connected to the internet.

- Agnew, J., P. Balduzzi, and A. Sunden (2003) "Portfolio Choice and Trading in a Large 401(k) Plan," *American Economic Review* 93, 193-215.
- Alba, J. and J. Hutchinson (1987) "Dimensions of Consumer Expertise," *Journal of Consumer Research* 12(4), 411-454.
- Andersen S., J. Campbell, K. Meisner-Nielsen and T. Ramadorai (2015) "Inattention and Inertia in Household Finance: Evidence from the Danish Mortgage Market," NBER Working Paper 21386.
- Atkinson, A. and F. Messy (2013), "Promoting Financial Inclusion Through Financial Education: OECD/INFE Evidence, Policies and Practice," OECD Working Papers on Finance, Insurance and Private Pensions, No. 34, OECD Publishing, Paris.
- Bazerman, M. and W. Samuelson (1983), "I Won the Auction But Don't Want the Prize," *Journal of Conflict Resolution* 27, 618-634.
- Baltagi, B. (2013) "Econometric Analysis of Panel Data", Wiley 5th edition.
- Barber, B. and T. Odean (2001) "Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment," *Quarterly Journal of Economics* 116(1), 261-292.
- Barberis, N. and R. Thaler (2003). "A Survey of Behavioral Finance," Handbook of the Economics of Finance, North-Holland.
- Baron-Cohen, S., Jolliffe, T., Mortimore, C. and M. Robertson (1997) "Another Advanced test of Theory of Mind: Evidence From Very High Functioning Adults With Autism or Asperger Syndrome," *Journal of Child Psychology and Psychiatry* 38, 813-822.
- Beauchamp, J., Cesarini, D., and M. Johannesson (2015) "The Psychometric Properties of Measures of Economic Risk Preferences," Working paper, New York University, New York.
- Benjamin, D., Brown S., and J. Shapiro, (2013) "Who is Behavioral? Cognitive Ability and Anomalous Preferences," Working paper, Harvard University.
- Biais, B., Hilton, D., Mazurier, K. and S. Pouget (2005) "Judgmental Overconfidence, Self-Monitoring and Trading Performance in an Experimental Financial Market," *The Review of Economic Studies* 72(2), 287-312.

Bossaerts, P. (2009) "What Decision Neuroscience Teaches us About Financial Decision Making," *Annual Review of Financial Economics* 1, 383-404.

Brañas-Garza, P., Kujal, P. and B. Lenkei (2015) "Cognitive Reflection Test: Whom, how, when," MPRA Paper 68049, University Library of Munich, Germany.

Brown, S. and K. Taylor (2014) "Household Finances and the 'Big Five' Personality traits," *Journal of Economic Psychology* 45I, 197-212.

Bruguier, A., Quartz, S. and P. Bossaerts (2010) "Exploring the Nature of 'Trader Intuition'," *Journal of Finance* 65, 1703-1723.

Chen, G., Kim, K., Nofsinger, J. and O. Rui (2007) "Trading Performance, Disposition Effect, Overconfidence, Representativeness Bias, and Experience of Emerging Market Investors," *Journal of Behavioral Decision Making* 20(4), 425-51.

Christelis, D., Tullio, J. and M. Padula (2010) "Cognitive Abilities and Portfolio Choice," *European Economic Review* 54, 18-38.

Campitelli, G. and M. Labollita (2010) "Correlations of Cognitive Reflection with Judgments and Choices," *Judgment and Decision Making* 5, 182-191.

Ciarrochi, J., Chan, A., and P. Caputi (2000) "A Critical Evaluation of the Emotional Intelligence Concept," *Personality and Individual Differences* 28, 539-561.

Cole, S. and G. Shastry (2009) "Smart Money: The Effect of Education, Cognitive Ability, and Financial Literacy on Financial Market Participation," Working paper, Harvard Business School, 09-071.

Conway, A., Kane, M. and R. Engle (2003) "Working Memory Capacity and its Relation to General Intelligence," *Trends in Cognitive Sciences* 7(12), 547-52.

Corgnet, B., Espín, A. and R. Hernan-Gonzalez (2015) "The Cognitive Basis of Social Behavior: Cognitive Reflection Overrides Antisocial but not Always Prosocial Motives," Economic Science Institute working papers, 15-04.

Corgnet, B., Hernan-Gonzalez, R., Kujal, P. and D. Porter (2014) "The Effect of Earned Versus House Money on Price Bubble Formation in Experimental Asset Markets," *Review of Finance* 19, 1455-1488.

Corgnet, B., DeSantis, M. and D. Porter (2015a) “On the Cognitive Foundations of Information Aggregation in Asset Markets: Reflective Learning & Market Efficiency,” Economic Science Institute working papers, 15-15.

Corgnet, B., M. DeSantis and D. Porter (2015b) “What Makes a Good Trader? On the Role of Quant Skills, Behavioral Biases and Intuition on Trading Performance” Economic Science Institute working paper 15-17.

Crack, T. (2014) “Heard on the Street: Quantitative Questions from Wall Street Job Interviews,” Dunedin, New Zealand.

Daniel, K., Hirshleifer, D., and A. Subrahmanyam (1998) “Investor Psychology and Security Market Under- and Overreactions,” *Journal of Finance* 53, 1839-1885.

De Martino, B., O’Doherty, J., Debajyoti, R., Bossaerts, P. and C. Camerer (2013) “In the Mind of The Market: Theory of Mind Biases Value Computation During Financial Bubbles,” *Neuron* 79, 1222-1231.

DeYoung, C. (2011) “Intelligence and Personality,” *The Cambridge Handbook of Intelligence*, Chapter 35, 711-737.

Dohmen, T. and A. Falk (2006) “Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender,” *American Economic Review* 10, 556-590.

Dohmen, T., Falk, A., Huffman, D. and U. Sunde (2010) “Are Risk Aversion and Impatience Related to Cognitive Ability?,” *American Economic Review* 100(3), 1238-60.

Easley D. and M. O’Hara (1987) “Price, Trade Size, and Information in Securities Markets,” *Journal of Financial Economics* 19(1), 69-90.

Ekins, G., Caceda, R., Capra, M. and G. Berns (2013) “You cannot gamble on others: Dissociable systems for strategic uncertainty and risk in the brain,” *Journal of Economic Behavior & Organization*, 94I, 222-233.

Eyster, E. and M. Rabin (2005) “Cursed Equilibrium”, *Econometrica* 73(5), 1623-1672.

Eyster, E., Rabin M. and D. Vayanos (2015) [“Financial Markets where Traders Neglect the Information Content in Prices”](#) CEPR Discussion Paper 10629.

- Fellner, G. and B. Maciejovsky (2007) "Risk Attitude and Market Behavior: Evidence From Experimental Asset Markets," *Journal of Economic Psychology* 28(3), 338-350.
- Fernandes, D., Lynch, J. and R. Netemeyer (2014) "Financial Literacy, Financial Education, and Downstream Financial Behaviors," *Management Science* 60(8), 1861-1883.
- Frederick, S. (2005) "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives* 19(4), 25-42.
- Frith, C. and U. Frith (1999) "Interacting Minds—A Biological Basis," *Science* 286, 1692-1695.
- Frydman, C., Barberis, N., Camerer, C., Bossaerts, P. and A. Rangel (2014) "Using Neural Data to Test a Theory of Investor Behavior: An Application to Realization Utility," *Journal of Finance* 69 (2), 907-946.
- Gaudecker, H. (2015) "How Does Household Portfolio Diversification Vary with Financial Literacy and Financial Advice?" *Journal of Finance* 70(2), 489-507.
- Golub, G. H., and C. F. Van Loan. 1996. *Matrix Computations*. 3rd ed. Baltimore: Johns Hopkins University Press.
- Greenspan, A. (2005) "The Importance of Financial Education Today," *Social Education* 69(2), 64-66.
- Glosten L. and P. Milgrom (1985) "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics* 14, 71-100.
- Grinblatt, M., Keloharju, M. and J. Linnainmaa (2012) "IQ, Trading Behavior, and Performance," *Journal of Financial Economics* 104(2), 339-362.
- Hampton, Alan N., Peter Bossaerts, and John P. O'Doherty, 2008, Neural correlates of mentalizing related computations during strategic interactions in humans, *Proceedings of the National Academy of Sciences* 105, 6741–6746.
- Holt, C. and S. Laury (2002) "Risk Aversion and Incentive Effects," *American Economic Review* 92(5), 1644-1655.
- Ibanez, A., Huepe, D., Gempp, R., Gutiérrez, V., Rivera-Rei, A. and M. Toledo (2013) "Empathy, Sex and Fluid Intelligence as Predictors of Theory of Mind," *Personality and Individual Differences* 5(5), 616-621.

Jaeggi, S., Studer-Luethi, B., Buschkuhl, M., Su, Y., Jonides, J. and W. Perrig (2010) “The Relationship Between N-Back Performance and Matrix Reasoning – Implications for Training and Transfer,” *Intelligence* 38(6), 625-635.

John, O., Donahue, E. and R. Kentle (1991) “The Big Five Inventory—Versions 4a and 54,” Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O., Naumann, L. and C. Soto (2008) “Paradigm Shift to the Integrative Big Five Trait Taxonomy: History, Measurement, And Conceptual Issues,” In O. John, R. Robins, and L. Pervin (Eds.), *Handbook of Personality: Theory and Research*, 3, 114-158. New York, NY: Guilford Press.

Jolliffe, I. (1995) “Rotation of Principal Components: Choice of Normalization Constraints,” *Journal of Applied Statistics* 22(1), 29-35.

Kahneman, D. (2011) “Thinking Fast and Slow,” New York: Farrar, Straus and Giroux.

Kezdi, G., and R. Willis (2003) “Who Becomes a Stockholder? Expectations, Subjective Uncertainty, and Asset Allocation,” Retirement Research Center Working paper, University of Michigan.

Kimbrough, E., Robalino, N. and A. Robson (2016) “Applying “Theory of Mind”: Theory and Experiments” SSRN working paper, <http://dx.doi.org/10.2139/ssrn.2791212>.

Larrick, R. (2004) “Debiasing,” Chapter 16 in Koehler DJ, Harvey N (eds.). *Blackwell Handbook of Judgment and Decision Making*. Blackwell publishing Ltd.

Lesage, E., Navarrete, G., and De Neys, W. (2013). Evolutionary modules and Bayesian facilitation: The role of general cognitive resources. *Thinking & Reasoning* 19, 27–53.

Lintner, J. (1969) “The Aggregation of Investor’s Diverse Judgments and Preferences in Purely Competitive Security Markets,” *Journal of Financial and Quantitative Analysis* 4(4), 347-400.

Logie, R., Gilhooly, K., and V. Wynn (1994) “Counting on Working Memory in Arithmetic Problem Solving,” *Memory & Cognition* 22, 395-410.

Loomis, C. (2013) “Tap Dancing to Work: Warren Buffett on Practically Everything, 1966-2013,” New York, NY: Penguin Group LLC.

Mackintosh, N. (2011) “History of Theories and Measurement of Intelligence,” The Cambridge Handbook of Intelligence, Chapter 1, 3-19.

McCabe, K., D. Houser, L. Ryan, V. Smith, and T. Trouard, 2001, A functional imaging study of cooperation in two-person reciprocal exchange, *Proceedings of the National Academy of Sciences* 98, 11832–11835.

Markowitz, H. (1952) “Portfolio Selection,” *Journal of Finance* 7, 77-91.

Mayer, J., Caruso, D. and P. Salovey (1999) “Emotional intelligence meets traditional standards for an intelligence,” *Intelligence* 27, 267-298.

Mayer, J., Salovey, P. and D. Caruso (2008) “Emotional Intelligence: New Ability or Eclectic Traits?” *American Psychologist* 63, 503-517.

Mayer, J., Salovey, P., Caruso, D. and L. Cherkasskiy (2011) “Emotional Intelligence,” The Cambridge Handbook of Intelligence, Chapter 26, 711-737.

Mishkin, F. (2008) “The Importance of Economic Education and Financial Literacy,” Presentation, Third National Summit on Economic Financial Literacy, February 27, Board of Governors of the Federal Reserve System, Washington, DC.

<http://www.federalreserve.gov/newsevents/speech/mishkin20080227a.htm>

Murphy, K. (1984) “The Wonderlic Personnel Test,” *Test Critiques*, 769-775. Keyser, D. J and Sweetland, R. C authors. Kansas City MO: Test Corporation of America.

Noussair, C. and S. Tucker (2014) “A Collection of Surveys on Market Experiments,” *Journal of Economic Surveys* 27(3), 395-397.

Noussair, C., Steven J. and Y. Xu (2014) “A Futures Market Reduces Bubbles But Allows Greater Profit for More Sophisticated Traders,” SSRN Working Paper.

Odean, T. (1998) “Are Investors Reluctant to Realize Their Losses?” *Journal of Finance* 53(3), 1175-1789.

Oechssler, J., Roider, A. and P. Schmitz (2009) “Cognitive Abilities and Behavioral Biases,” *Journal of Economic Behavior & Organization* 72(1), 147-152.

- Parsons, S., Marcinkiewicz M., Niu, J. and S. Phelps (2008) “Everything you Wanted to Know About Double Auctions, but Were Afraid to (bid or) ask,” Working Paper, Department of Computer and Information Science, Brooklyn College.
- Payzan-LeNestour, E. and P. Bossaerts (2015) “Learning About Unstable, Publicly Unobservable Payoffs,” *Review of Financial Studies* 28(7), 1874-1913.
- Plott, C. and S. Sunder (1988) “Rational Expectations and the Aggregation of Diverse Information in Laboratory Security Markets,” *Econometrica* 56, 1085-1118.
- Pouget, S. 2007. “Adaptive Traders and the Design of Financial Markets,” *Journal of Finance* 62, 2835-2863.
- Rabin, M. (2002) “Inference by Believers in the Law of Small Numbers,” *The Quarterly Journal of Economics* 117(3), 775-816.
- Ramsden, S., Richardson, F., Josse, G., Thomas, M., Ellis, C., Shakeshaft, C., Seghier, M. and C. Price (2011) “Verbal and Nonverbal Intelligence Changes in the Teenage Brain,” *Nature* 479 (7371), 113-116.
- Raven, J. (1941) “Standardization of Progressive Matrices,” *British Journal of Medical Psychology* 19(1), 137-150.
- Russo, J. and P. Schoemaker (1992) “Managing Over-Confidence,” *Sloan Management Review*, 33, 7-17.
- Sharpe, W. (1964) “Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk,” *Journal of Finance* 19(3), 425-442.
- Shefrin, H. (2007) “Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing,” New York: Oxford University Press.
- Sirota, M., Juanchich, M., and Hagmayer, Y. (2014). Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. *Psychonomic Bulletin & Review*, 21, 198–204.
- Snyder, M. and S. Gangestad (1986) “On the Nature of Self-Monitoring: Matters of Assessment, Matters of Validity,” *Journal of Personality and Social Psychology* 51:125-139.

- Stanovich, K. (2009a) “What Intelligence Tests Miss: The Psychology of Rational Thought,” Yale University Press.
- Stanovich, K. (2009b) “Decision Making and rationality in the Modern World,” Oxford University Press.
- Stanovich, K. (2016) “The rationality Quotient (RQ): Toward a Test of Rational Thinking,” MIT Press.
- Svenson, O. (1981) “Are we all Less Risky and More Skillful Than our Fellow Drivers?,” *Acta Psychologica* 47, 143-148.
- Thaler, R. H. (1988) “Anomalies: The Winner’s Curse,” *Journal of Economic Perspectives* 2(1), 191-202.
- Thaler, R. H. (1991) “The Winner’s Curse: Paradoxes and Anomalies of Economic Life,” Free Press, 1991.
- Thaler, R. H. (1993) “Advances in Behavioral Finance,” New York: Russell Sage Foundation.
- Thaler, R. H. (2005) “Advances in Behavioral Finance,” Volume II, Princeton: Princeton University Press.
- Thoma, V., White, E., Panigrahi, A., Strowger, V. and I. Anderson (2015) “Good Thinking or Gut Feeling? Cognitive Reflection and Intuition in Traders, Bankers and Financial Non-Experts,” *PloS One* 10(4).
- Toplak, M., West, R., and K. Stanovich (2011) “The Cognitive Reflection Test as a Predictor of Performance on Heuristics and Biases Tasks,” *Memory & Cognition* 39, 1275-1289.
- Toplak, M., West, R., and K. Stanovich (2014) “Assessing Miserly Information Processing: An Expansion of the Cognitive Reflection Test,” *Thinking & Reasoning* 20(2), 147-168.
- Tversky, A. and D. Kahneman (1974) “Judgment under Uncertainty: Heuristics and Biases,” *Science* 185, 1124-1131.
- US Congress (2010) Dodd–Frank Wall Street Reform and Consumer Protection Act, HR 4173.ENR, 111th Cong., 2nd sess. Accessed May 31, 2013.
<http://www.gpo.gov/fdsys/pkg/BILLS-111hr4173enr/pdf/BILLS-111hr4173enr.pdf>

Zhou, X. (2008) “A Practical Guide to Quantitative Finance Interviews,” CreateSpace Independent Publishing Platform; 14th edition.

Internet Appendix

Appendix A. Survey Description

One-hour follow-up lab survey

Extended cognitive reflection test (CRT):

Taken from Frederick (2005):

- (1) A bat and a ball cost \$1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? ____ cents
[Correct answer: 5 cents; intuitive answer: 10 cents]
- (2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes
[Correct answer: 5 minutes; intuitive answer: 100 minutes]
- (3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____ days
[Correct answer: 47 days; intuitive answer: 24 days]

Taken from Toplack et al. (2014):

- (4) If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together? ____ days
[correct answer: 4 days; intuitive answer: 9]
- (5) Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class? ____ students
[correct answer: 29 students; intuitive answer: 30]
- (6) A man buys a pig for \$60, sells it for \$70, buys it back for \$80, and sells it finally for \$90. How much has he made? ____ dollars
[correct answer: \$20; intuitive answer: \$10]

- (7) Simon decided to invest \$8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he began, c. has lost money
[correct answer: c; intuitive response: b]

Table A.1. Distribution of CRT scores

CRT score	% of subjects
0	8.82
1	18.62
2	19.11
3	17.65
4	17.15
5	6.85
6	6.85
7	4.95
Mean	2.88
Standard Deviation	1.89

For each of the following tests, CRT, eye gaze and Raven, we asked participants to rank themselves as follows: “*Out of the last ten people that took this test, how would you rank yourself: First, Second, ..., Tenth.*”

Eye gaze test

This is an example of one of the 36 questions in the test of Baron-Cohen (1997):



Figure A.1: Example of an eye gaze test question

Table A.2. Distribution of eye gaze test scores

Eye gaze score	0-9	10-15	16-20	21-25	26-30	>30	Mean	Stand. Dev.
% of subjects	4.19	0.00	5.39	21.56	55.68	13.18	25.86	6.24

Raven Test

We utilized the Raven progressive matrices test (Raven, 1941) as a general measure of intelligence (Mackintosh, 2011). Specifically, we utilized the odd number of the last three series of matrices (Jaeggi et al. 2010). The duration of the test was 10 minutes. The number of matrices correctly

solved in the Raven test is a conventional measure of cognitive ability. Measures of general intelligence are commonly related to working memory capacity, which refers to the short-term holding and manipulation of information (Conway, Kane and Engle, 2003). The following is an example of a Raven question:

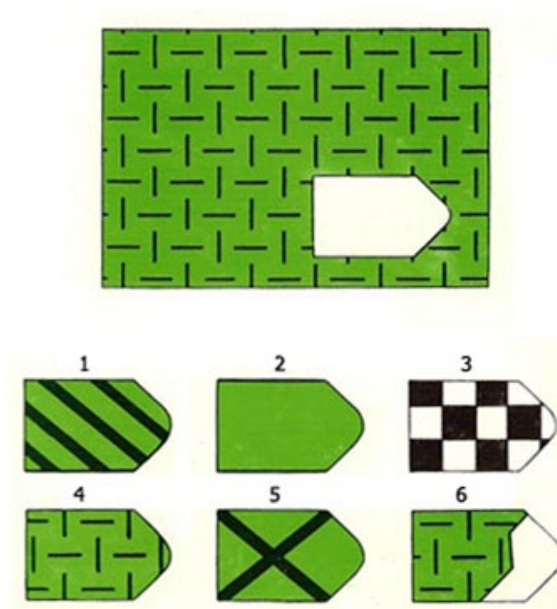


Figure A.2: Example of a Raven Test question

Table A.3. Distribution of Raven scores (Study 1)

Raven score	<10	10	11	12	13	14	15	16	17	18	Mean	Standard Deviation

% of subjects	6.37	1.47	7.35	9.80	18.63	15.69	18.63	12.75	6.86	2.45	13.69	2.42
---------------	------	------	------	------	-------	-------	-------	-------	------	------	-------	------

Risk attitudes

The risk elicitation task was taken from Holt and Laury (2002). The instructions were as follows.

In this part of the experiment you will be asked to make a series of choices in decision problems. How much you receive will depend partly on chance and partly on the choices you make. The decision problems are not designed to test you. What we want to know is what choices you would make in them. The only right answer is what you really would choose. For each line in the table on the right, please state whether you prefer option A or option B. Notice that there are a total of 10 lines in the table but just one line will be randomly selected for payment. Each line is equally likely to be chosen, so you should pay equal attention to the choice you make in every line. At the end of the experiment, a number between 1 and 10 will be randomly selected by the computer. This number determines which line is going to be paid. Your earnings for the selected line depend on which option you chose in that line: option A or option B. To determine your earnings, then, a second number between 1 and 10 will be randomly selected by the computer. This number is then compared with the numbers in the line and option selected (see the table below).

If the first number selects line 4:

* If you selected option A you earn \$2.00 if the second number is 1, 2, 3 or 4 and you earn \$1.60 otherwise.

* If you selected option B you earn \$3.85 if the second number is 1, 2, 3 or 4 and you earn \$0.10 otherwise.

Table A.4. The ten binary lottery choices (Holt and Laury, 2002).

Line	OPTION A	OPTION B
1	1/10 of \$2.00, 9/10 of \$1.60	1/10 of \$3.85, 9/10 of \$0.10

2	2/10 of \$2.00, 8/10 of \$1.60	2/10 of \$3.85, 8/10 of \$0.10
3	3/10 of \$2.00, 7/10 of \$1.60	3/10 of \$3.85, 7/10 of \$0.10
4	4/10 of \$2.00, 6/10 of \$1.60	4/10 of \$3.85, 6/10 of \$0.10
5	5/10 of \$2.00, 5/10 of \$1.60	5/10 of \$3.85, 5/10 of \$0.10
6	6/10 of \$2.00, 4/10 of \$1.60	6/10 of \$3.85, 4/10 of \$0.10
7	7/10 of \$2.00, 3/10 of \$1.60	7/10 of \$3.85, 3/10 of \$0.10
8	8/10 of \$2.00, 2/10 of \$1.60	8/10 of \$3.85, 2/10 of \$0.10
9	9/10 of \$2.00, 1/10 of \$1.60	9/10 of \$3.85, 1/10 of \$0.10
10	10/10 of \$2.00, 0/10 of \$1.60	10/10 of \$3.85, 0/10 of \$0.10

Table A.5. Number of safe choices in the Holt and Laury task

Number of safe choices	0	1	2	3	4	5	6	7	8	9	10	Mean	Standard Deviation
% of subjects	4.19	2.40	1.20	6.59	19.75	17.36	26.35	16.16	2.40	1.80	1.80	4.95	2.02

Personality traits

The Big Five Inventory (John, Donahue and Kentle , 1991 and John, Naumann and Soto, 2008).

Table A.6. Big five traits descriptive statistics

Variable	α	Mean	Std. Dev.
Openness	.82	37.6	6.6

Conscientiousness	.80	32.2	5.4
Extraversion	.84	27.3	5.9
Agreeableness	.81	33.8	5.3
Neuroticism	.84	23.4	6.2

The following demographic information was collected: gender, age, school attended and GPA. Additional measures, such as the Wonderlic verbal test and adding skills (see Corgnet, DeSantis and Porter, 2015b), were also collected. Inclusion of these variables in our robustness checks analyses (see Appendix F1) does not affect the qualitative nature of our results.

End-of-experiment survey

Financial literacy

The scale was taken from Fernandes, Lynch and Netemeyer (2014).

- 1) Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, would you be able to buy:
 - ☐ more than today with the money in this account
 - ☐ exactly the same as today with the money in this account
 - ☐ *less than today with the money in this account*
 - ☐ Don't know
 - ☐ Refuse to answer

- 2) Do you think that the following statement is true or false? "Bonds are normally riskier than stocks."
 - ☐ True
 - ☐ *False*

- ☐ Don't know
- ☐ Refuse to answer

3) Considering a long time period (for example 10 or 20 years), which asset described below normally gives the highest return?

- ☐ savings accounts
- ☐ *stocks*
- ☐ bonds
- ☐ Don't know
- ☐ Refuse to answer

4) Normally, which asset described below displays the highest fluctuations over time?

- ☐ savings accounts
- ☐ *stocks*
- ☐ bonds
- ☐ Don't know
- ☐ Refuse to answer

5) When an investor spreads his money among different assets, does the risk of losing a lot of money:

- ☐ increase
- ☐ *decrease*
- ☐ stay the same
- ☐ Don't know
- ☐ Refuse to answer

6) Do you think that the following statement is true or false? "If you were to invest \$1000 in a stock mutual fund, it would be possible to have less than \$1000 when you withdraw your money."

- ☐ *True*

- ☐ False
- ☐ Don't know
- ☐ Refuse to answer

7) Do you think that the following statement is true or false? "A stock mutual fund combines the money of many investors to buy a variety of stocks."

- ☐ *True*
- ☐ False
- ☐ Don't know
- ☐ Refuse to answer

8) Do you think that the following statement is true or false? "After age 70 ½, you have to withdraw at least some money from your 401(k) plan or IRA."

- ☐ *True*
- ☐ False
- ☐ It depends on the type of IRA and/or 401(k) plan
- ☐ Don't know
- ☐ Refuse to answer

9) Do you think that the following statement is true or false? "A 15-year mortgage typically requires higher monthly payments than a 30-year mortgage, but the total interest paid over the life of the loan will be less."

- ☐ *True*
- ☐ False
- ☐ Don't know
- ☐ Refuse to answer

10) Suppose you had \$100 in a savings account and the interest rate is 20% per year and you never withdraw money or interest payments. After 5 years, how much would you have on this account in total?

- ☐ *More than \$200*
- ☐ Exactly \$200
- ☐ Less than \$200
- ☐ Don't know
- ☐ Refuse to answer

11) Which of the following statements is correct?

- ☐ Once one invests in a mutual fund, one cannot withdraw the money in the first year
- ☐ *Mutual funds can invest in several assets, for example invest in both stocks and bonds*
- ☐ Mutual funds pay a guaranteed rate of return which depends on their past performance
- ☐ None of the above
- ☐ Don't know
- ☐ Refuse to answer

12) Which of the following statements is correct? If somebody buys a bond of firm B:

- ☐ He owns a part of firm B
- ☐ *He has lent money to firm B*
- ☐ He is liable for firm B's debts
- ☐ None of the above
- ☐ Don't know
- ☐ Refuse to answer

13) Suppose you owe \$3,000 on your credit card. You pay a minimum payment of \$30 each month. At an Annual Percentage Rate of 12% (or 1% per month), how many years would it take to eliminate your credit card debt if you made no additional new charges?

- ☐ less than 5 years

- ☐ between 5 and 10 years
- ☐ between 10 and 15 years
- ☐ *never*
- ☐ Don't know
- ☐ Refuse to answer

Table A.7. Distribution of financial literacy scores (Study 1)

Financial literacy score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	Mean	Standard Deviation
% of subjects	2.45	3.43	2.94	3.43	8.82	14.71	17.16	14.71	9.80	9.80	6.37	3.92	1.96	0.50	6.38	2.69

Self-monitoring scale

This scale is taken from Snyder and Gangestad (1986) and it was used by Biais et al. (2005). Subjects have to answer whether each of 18 statements accurately describes their social behavior. We obtained a Cronbach alpha (0.71) similar to the one reported by Biais et al. (2005) (0.70). Subjects had 4 minutes to complete the test.

For each of the following questions, we code 1 if the answer reflects self-monitoring and 0 otherwise. Our measure of the degree to which the participant is a self-monitor is the percentage of answers coded with a 1.

Answer each Question by TRUE or FALSE:

I find it hard to imitate the behavior of other people.

At parties and social gatherings, I do not attempt to do or say things that others will like

I can only argue for ideas, which I already believe.

I can make impromptu speeches even on topics about which I have almost no information.

I guess I put on a show to impress or entertain others.

I would probably make a good actor.
 In a group of people I am rarely the center of attention.
 In different situations and with different people, I often act like very different persons.
 I am not particularly good at making other people like me.
 I'm not always the person I appear to be.
 I would not change my opinions (or the way I do things) in order to please someone or win their favor.
 I have considered being an entertainer.
 I have never been good at games like charades or improvisations.
 I have trouble changing my behavior to suit different people and different situations.
 At a party I let others keep the jokes and stories going.
 I feel a bit awkward in public and do not show up quite as well as I should.
 I can look anyone in the eyes and tell a lie with a straight face.
 I may deceive people by being friendly when I really dislike them.

Table A.8. Distribution of self-monitoring scores (Study 1)

Self-monitoring score	0-3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean	Standard Deviation
% of subjects	6.37	8.34	6.87	7.35	8.34	8.34	4.90	10.78	7.35	7.84	10.29	4.41	7.35	1.47	9.09	3.76

Bayesian updating test (collected in 5 high stakes sessions, n=60)

Following Charness and Levin (2009) we use the following test.

Consider two machines placed in two sides of large production hall, left side = L and right side = R. The two machines produce rings, good ones and bad ones. Each ring that comes from the left machine, L, has a 50 percent chance of being a good ring and a 50 percent chance of being a bad ring. Each ring that comes from the right machine, R, has a 75 percent chance of being a good ring and a 25 percent chance of being a bad ring. In each of the following four questions, you will observe some ring(s) that are the output of one of the two machines. After the information is given, you are asked to mark one (and only one) of the probabilities that you think is closest to the true probability that the L = left machine produced the ring(s), given your observations.

Question 1. You observe one good ring. What is your best assessment of the probability that the left machine produced this ring (select the right answer)?

100%	90%	80%	70%	60%	50%	40%	30%	20%	10%	0%
------	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

Question 2. You observe one bad ring. What is your best assessment of the probability that the left machine produced this ring?

100%	90%	80%	70%	60%	50%	40%	30%	20%	10%	0%
------	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

Question 3. You observe six rings, and all six are bad. What is your best assessment of the probability that the left machine produced these rings?

100%	90%	80%	70%	60%	50%	40%	30%	20%	10%	0%
------	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

Question 4. You observe six rings, of which three are good and three are bad. What is your best assessment of the probability that the left machine produced these rings?

100%	90%	80%	70%	60%	50%	40%	30%	20%	10%	0%
------	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

Each question is scored by taking the absolute difference between a subject's answer and the correct answer for each of the four questions. A subject's final score is minus the sum of each question score.

Table A.9. Distribution Bayesian updating scores

Bayes score	[0-0.25)	[0.25-0.50)	[0.50-0.75)	[0.75-1.00)	[1.00-1.25)	[1.25-1.50)	≥ 1.50	Mean	Standard Deviation
% of subjects	0	3.3	21.7	38.3	11.7	15	10	1.03	0.433

We report a Cronbach alpha of 0.67 for this test.

The Raven test and the CRT were also collected as in the one-hour lab survey.

*Question on session CRT ranking (collected in 4 high stakes sessions, $n=48$)*³³

Participants were asked the following question:

All of the participants in this experiment have taken a previous 7-question cognitive test. One example of a question was “A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?” Out of the 12 people who participated in the current market experiment, where would you rank yourself in terms of this cognitive test? (1 means FIRST and 12 means LAST?)

Table A.10. Distribution of *perceived* CRT ranks in a given session

Rank	1	2	3	4	5	6	7	8	9	10	11	Mean	Standard Deviation
% of subjects	29.17	12.50	6.25	8.33	16.67	10.42	6.25	2.08	4.17	2.08	2.08	3.94	2.74

³³ This question was asked to all 5 high stakes sessions but a technical issue made it impossible to collect the information for one session.

Appendix B. Model and hypotheses

Appendix B1. Simulation procedure

Reflective traders in our model use Bayes' rule to update their belief of the asset's true value after each market event. These events can be: (1) the current best bid is traded, (2) the current best ask is traded and (3) both the bid and the ask are (strictly) improved (i.e., no trade occurs). The updating rules for each event are described in this section. Consider the following notation:

- v : value of the asset
- τ : traders' behavioral type where $\tau \in \{R_P, R_{NP}, NR\}$ with R_P (R_{NP}) [NR] signifying a *perceptive reflective* (*non-perceptive reflective*) [*non-reflective*] trader
- h : hint given to a trader; $h \in \{\text{"Not 50"}, \text{"Not 240"}, \text{"Not 490"}\}$
- (τ, h) : trader *group* defined as a combination of a behavioral type and hint, e.g., $(\tau, h) = (\text{perceptive reflective}, \text{"Not 50"})$
- α : proportion of *reflective* traders in the market
- α^τ : represents the belief traders of type τ hold for α ; note that $\alpha^{R_P} = \alpha$ for *perceptive reflective* traders while $\alpha^{R_{NP}} \neq \alpha$ for *non-perceptive reflective* traders
- b_{n-1} : current best bid at the time of event n (a default bid of 50 is always present)³⁴
- a_{n-1} : current best ask at the time of event n (assume default ask of 490 is always present)
- b_n : newly submitted bid corresponding to event n (improves upon b_{n-1})
- a_n : newly submitted ask corresponding to event n (improves upon a_{n-1})
- $T(o)$: the current best offer, b_{n-1} or a_{n-1} , is traded
- $NT(o)$: the current best offer, b_{n-1} or a_{n-1} , is not traded; so, the current best bid (ask) is improved, i.e. $b_n > b_{n-1}$ ($a_n < a_{n-1}$)

³⁴ Note that the current best bid (ask) at the time of event n may not correspond to the $n - 1^{st}$ bid (ask) submitted to the market due to events such as trades that do not update the bid (ask). Consider, for example, the bids in Figure 1. The initial default bid is $b_0 = 50$. The first event updates this bid to $b_1 = 200$. The second event is a transaction at the price $p_2 = 360$. The third event corresponds to the submission of a new bid, $b_3 = 200$, to the market. This new bid updates the current best bid, which is 200. Thus, in a situation like this we implicitly assume $b_2 = b_1 = 200$.

$e(n)$: represents the n^{th} event; possible events are: (1) $\{T(b_{n-1})\}$, (2) $\{T(a_{n-1})\}$ and (3) $\{NT(b_{n-1}) \cap NT(a_{n-1})\}$, i.e. the current bid-ask spread, (b_{n-1}, a_{n-1}) , is improved to (b_n, a_n)

$\mu_{n,h}^\tau$: belief of trader of behavioral type τ with the hint h after event n

$Prior_h^v(n-1)$: probability assigned to the true value, v , by the trader with the hint h after the $(n-1)^{st}$ event

$u_{X_r}(l_{X_r})$: upper (lower) bound of region X_r where $r \in \{b_{n-1}, a_{n-1}, b_n, a_n\}$ represents the current or newly submitted request to buy or sell. Regions are defined by the current beliefs of all trader *groups* and also depend upon the assumed asset value.³⁵ Suppose that a *reflective* trader believes the asset's value to be 50. Then four *groups* of traders exist: *reflective* with the hint "Not 240", *reflective* with the hint "Not 490", *non-reflective* with the hint "Not 240" and *non-reflective* with the hint "Not 490".³⁶ The beliefs of these four *groups* of traders are used to discretize the interval $[50, 490]$ into at most 5 regions (less if the belief of one trader type coincides with that of another type or with one of the possible asset values, 240 or 490).

$$\rho_{buy}^{X_a} = P[a \text{ trader buys in } X_a] = P[\text{ask is accepted in } X_a]$$

$$\rho_{sell}^{X_b} = P[a \text{ trader sells in } X_b] = P[\text{bid is accepted in } X_b]$$

We restrict bids and asks (and hence prices) to be in the interval $[50, 490]$ which has the minimum possible value of the asset as a lower bound and the maximum possible value of the asset as an upper bound. As noted in Section 4, traders make decisions to buy (sell) depending upon their belief of the asset's value. If their belief is greater (less) than the current best ask (bid), then they will buy (sell) the asset provided they enough cash (at least one share). Otherwise, they will submit a new bid and ask that strictly improves the current bid-ask spread. The new bid is drawn from a uniform distribution over the interval $[\max(b_{n-1}, \mu_{n-1,h}^\tau), \min(a_{n-1}, \mu_{n-1,h}^\tau)]$.

³⁵ See, for example, Figure B2.1.

³⁶ The *reflective* trader only considers four *groups* as she assumes other *reflective* traders share her belief of the proportion of *reflective* traders in the market. Thus, if her type is *perceptive reflective* (*non-perceptive reflective*), she assumes other *reflective* traders are also *perceptive* (*non-perceptive*).

Without loss of generality suppose that after the n^{th} event, the trader with the hint “Not 240” updates her belief as follows³⁷:

$$\begin{aligned}
\mu_{n, \text{Not } 240}^R &= \\
&= E[v | \mu_{n-1, \text{Not } 240}^R, \{e(n)\}] \\
&= P[v = 50 | \{e(n)\}] \times 50 + P[v = 490 | \{e(n)\}] \times 490 \\
&= \frac{P[\{e(n)\} | v = 50] \text{Prior}_{240}^{50}(n-1)}{P[\{e(n)\}]} \times 50 + \frac{P[\{e(n)\} | v = 490] \text{Prior}_{240}^{490}(n-1)}{P[\{e(n)\}]} \times 490
\end{aligned}$$

where

$$P[\{e(n)\}] = P[\{e(n)\} | v = 50] \times \text{Prior}_{240}^{50}(n-1) + P[\{e(n)\} | v = 490] \times \text{Prior}_{240}^{490}(n-1).$$

We need to calculate $P[\{e(n)\} | v]$. Let α^τ represent this trader’s belief of the asset value where $\alpha^{RP} = \alpha$ for *perceptive reflective* traders while $\alpha^{RNP} \neq \alpha$ for *non-perceptive reflective* traders. There are three cases to consider.

Case 1: *the current best bid is traded*, $\{e(n)\} = \{T(b_{n-1})\}$

Suppose event n corresponds to the acceptance of the current best bid, b_{n-1} . Then the conditional probability that event $\{T(b_{n-1})\}$ occurred, given the true value of the asset is v , is

$$P[\{e(n)\} | v] = P[\{T(b_{n-1})\} | v] = \rho_{sell}^{X_{b_{n-1}}}$$

To calculate the probability of selling in region $X_{b_{n-1}}$ we consider the belief of each trader group. Suppose $v = 50$ in the above equation and set $\rho_{sell}^{b_{n-1}}$ equal to zero. If the belief of *perceptive reflective* (*non-perceptive reflective*) traders with the hint “Not 240” (“Not 490”) is

³⁷ Note that when a *reflective* trader calculates a probability, it is assumed that the probability is determined given the trader’s hint.

less than or equal to the lower bound of the region, then add $\alpha^\tau/2$ to $\rho_{sell}^{X_{b_{n-1}}}$. Similarly, when considering whether *non-reflective* traders would sell in the region, we add $(1 - \alpha^\tau)/2$ for each group.

Case 2: *the current best ask is traded*, $\{e(n)\} := \{T(a_{n-1})\}$

Suppose the current best ask, a_{n-1} , is accepted. Then the conditional probability that event $\{T(a_{n-1})\}$ occurred, given the true value of the asset is v , is:

$$P[\{e(n)\}|v] = P[\{T(a_{n-1})\}|v] = \rho_{buy}^{X_{a_{n-1}}}$$

In this case we determine whether the belief of *reflective* (*non-reflective*) traders with the hint “Not 240” (“Not 490”) is greater than or equal to the upper bound of the region and add $\alpha^\tau/2$ ($[1 - \alpha^\tau]/2$) to $\rho_{buy}^{X_{a_{n-1}}}$.

Case 3: *both the current best bid and ask are improved*, $\{e(n)\} := \{NT(b_{n-1}) \cap NT(a_{n-1})\}$

Suppose neither the current best bid nor the current best ask is traded. Then both the bid and the ask are updated from the current best bid, b_{n-1} , and ask, a_{n-1} , to the new (improved) best bid, b_n , and ask, a_n . Then the conditional probability that event $\{e(n)\} = \{NT(b_{n-1}) \cap NT(a_{n-1})\}$ occurred, given the true value of the asset is v , is:

$$\begin{aligned} P[\{e(n)\}|v] &= P[\{NT(b_{n-1}) \cap NT(a_{n-1})\} | v, b_{n-1}, a_{n-1}] \\ &= P[NT(b_{n-1}) | v, b_{n-1}, a_{n-1}] & [i] \\ &\times P[NT(a_{n-1}) | v, b_{n-1}, a_{n-1}, NT(b_{n-1})] & [ii] \\ &\times P[b_n \in X_{b_n} \cap a_n \in X_{a_n} | v, b_{n-1}, a_{n-1}] & [iii] \end{aligned}$$

Consider each term in this product separately.

[i] The first term is equivalent to one minus the probability that the current best bid, b_{n-1} , was not traded. This is equal to one minus the probability that traders are willing to sell in the region $X_{b_{n-1}}$. Thus, we have:

$$P[\text{NT}(b_{n-1}) | v, b_{n-1}, a_{n-1}] = (1 - \rho_{\text{sell}}^{X_{b_{n-1}}}).$$

[ii] The second term may be rewritten as:

$$\begin{aligned} & P[\text{NT}(a_{n-1}) | v, b_{n-1}, a_{n-1}, \text{NT}(b_{n-1})] \\ &= 1 - P[\text{T}(a_{n-1}) | v, \text{NT}(b_{n-1})] \\ &= 1 - P[a \text{ trader buys in } X_{a_{n-1}} | v, a \text{ trader buys in } X_{b_{n-1}}] \\ &= 1 - P[a \text{ trader buys in } X_{b_{n-1}} | v, a \text{ trader buys in } X_{a_{n-1}}] \times \frac{P[a \text{ trader buys in } X_{a_{n-1}}]}{P[a \text{ trader buys in } X_{b_{n-1}}]} \\ &= 1 - \rho_{\text{buy}}^{X_{a_{n-1}}} / \rho_{\text{buy}}^{X_{b_{n-1}}} \end{aligned}$$

where we make use of Bayes' rule as well as the fact that since $X_{a_{n-1}}$ must lie to the right of (or possibly coincide with) $X_{b_{n-1}}$ on the number line, $P[a \text{ trader buys in } X_{b_{n-1}} | v, a \text{ trader buys in } X_{a_{n-1}}] = 1$.

[iii] Finally, the third term may be expressed as:

$$\begin{aligned} & P[b_n \in X_{b_n} \cap a_n \in X_{a_n} | v, b_{n-1}, a_{n-1}] = P[b_n \in X_{b_n} | v, b_{n-1}, a_{n-1}] \times P[a_n \in X_{a_n} | v, b_{n-1}, a_{n-1}, b_n] \\ &= \left[\sum_{\substack{(\tau, h); \\ h \neq \text{"Not } v"}}} \frac{\gamma}{2} \left\{ \frac{\min(u_{X_{b_n}}, a_{n-1}) - \max(l_{X_{b_n}}, b_{n-1})}{\min(\mu_{n-1, h}^\tau, a_{n-1}) - \max(50, b_{n-1})} \right\} \right] \\ &\quad \times \left[\sum_{\substack{(\tau, h); \\ h \neq \text{"Not } v"}}} \frac{\gamma}{2} \left\{ \frac{\min(u_{X_{a_n}}, a_{n-1}) - \max(l_{X_{a_n}}, b_n)}{\min(490, a_{n-1}) - \max(\mu_{n-1, h}^\tau, b_n)} \right\} \right] \end{aligned}$$

where the sums are taken over the four trader *groups* that did not receive the hint “Not v .” If the true value is assumed to be 50, then the four *groups* are: *reflective* traders with the hint “Not

240”, *reflective* traders with the hint “Not 490”, *non-reflective* traders with the hint “Not 240” and *non-reflective* traders with the hint “Not 490.” The variable γ is set to α^τ for *reflective* trader groups and $(1 - \alpha^\tau)$ for *non-reflective* trader groups (see Table B1.1).³⁸

Table B1.1. Proportion of traders from the perspective of a *reflective* trader with hint “Not 240”

	Asset Value = 50		Asset Value = 490	
	Hint, h		Hint, h	
Behavioral type, τ	“Not 240”	“Not 490”	“Not 50”	“Not 240”
<i>Perceptive reflective</i> , R_P	$\alpha^{R_P}/4$	$\alpha^{R_P}/4$	$\alpha^{R_P}/4$	$\alpha^{R_P}/4$
<i>Non-perceptive reflective</i> , R_{NP}	$\alpha^{R_{NP}}/4$	$\alpha^{R_{NP}}/4$	$\alpha^{R_{NP}}/4$	$\alpha^{R_{NP}}/4$
<i>Non-reflective</i> , NR	$(1 - \alpha^\tau)/2$	$(1 - \alpha^\tau)/2$	$(1 - \alpha^\tau)/2$	$(1 - \alpha^\tau)/2$

As noted above, $\alpha^\tau = \alpha^{R_P} = \alpha$ for *perceptive reflective* traders while $\alpha^\tau = \alpha^{R_{NP}} \neq \alpha$ for *non-perceptive reflective* traders. Each cell corresponds to a trader group. There are nine total trader groups, though only six exist given the true value of the asset.

Thus, we have

$$\begin{aligned}
 P[\{e(n)\}|v] &= \left(1 - \rho_{sell}^{X_{b_{n-1}}}\right) \times \left(1 - \frac{\rho_{buy}^{X_{a_{n-1}}}}{\rho_{buy}^{X_{b_{n-1}}}}\right) \times \left[\sum_{\substack{(\tau, h); \\ h \neq \text{"Not } v"}} \frac{\gamma}{2} \left\{ \frac{\min(u_{X_{b_n}}, a_{n-1}) - \max(l_{X_{b_n}}, b_{n-1})}{\min(\mu_{n-1, h}^\tau, a_{n-1}) - \max(50, b_{n-1})} \right\} \right] \\
 &\quad \times \left[\sum_{\substack{(\tau, h); \\ h \neq \text{"Not } v"}} \frac{\gamma}{2} \left\{ \frac{\min(u_{X_{a_n}}, a_{n-1}) - \max(l_{X_{a_n}}, b_n)}{\min(490, a_{n-1}) - \max(\mu_{n-1, h}^\tau, b_n)} \right\} \right]
 \end{aligned}$$

³⁸ If there exists an odd number of traders, then the assumption is that traders are split as evenly as possible across the different types so that, at most, there exists one more trader of a certain type. For example, if $\alpha = 5/12$, then there might be two *perceptive reflective* traders and three *non-perceptive reflective* traders. In addition, if the asset value is assumed to be 50, then one of the *non-perceptive reflective* traders might receive the clue “Not 240”, while the other two receive the clue “Not 490”.

when the n^{th} event is an improvement to the bid-ask spread.³⁹ As previously noted, *reflective* traders assume other *reflective* traders share their assessment of the proportion of *reflective* traders in the market. Thus, the belief updates for *perceptive reflective* and *non-perceptive reflective* traders may differ as demonstrated in the example in the main text.

Appendix B2. Simulation example

We illustrate our model by providing a more general example than the one provided in the main text (Section 4.3) where the market is populated by both *reflective* and *non-reflective* traders. In particular, we consider the case in which there are four *reflective* traders in the market. Thus, $\alpha = 4/12$, while *non-perceptive reflective* traders assume the proportion of reflective traders in the market is $\alpha^{RNP} = 9/12$. The true value of the asset is set to 240. As such, the exogenous parameters for this simulation example are: $(v, \alpha, \alpha^{RNP}) = (240, 4/12, 9/12)$.

Based on their prior information (“Not 50”, “Not 490” or “Not 240”), *reflective* traders’ initial beliefs regarding the value of the asset are: $\mu_{0,Not\ 50}^R = 316.9$, $\mu_{0,Not\ 490}^R = 156.9$ and $\mu_{0,Not\ 240}^R = 210$, where $\mu_{0,h}^R$ represents the belief of *reflective* traders who received the hint h prior to the occurrence of any order.⁴⁰ *Non-reflective* traders’ prior beliefs are the same as those of *reflective* traders: $\mu_{0,Not\ 50}^{NR} = 316.9$, $\mu_{0,Not\ 490}^{NR} = 156.9$ and $\mu_{0,Not\ 240}^{NR} = 210$.

First market event: bid-ask spread {76, 332}

Let us consider traders with the hint “Not 50”.⁴¹ Suppose the first trader submits an initial bid and ask spread of $\{b_1 = 76, a_1 = 332\}$. Both *reflective* traders with the “Not 50” hint will update their belief to 327.71 by applying the following Bayesian formula:

³⁹ Strictly speaking, there exist cases in which only the bid or the ask is improved. However, we assume traders do not update their beliefs on these unlikely events. These cases occur, for example, when a trader wants to update both the bid and the ask, but does not have the requisite amount of cash/shares to cover the updated bid/ask as well as any outstanding bids/asks the trader might have.

⁴⁰ As the true value of the asset is assumed to be 240, there are no traders with the hint “Not 240” in the market. However, this fact is not known by the *reflective* traders who must therefore act as if traders with the hint “Not 240” could potentially exist.

⁴¹ The updating rules for other traders and other possible asset values are similar.

$$\mu_{n, \text{Not } 50}^R = \frac{P[\{e(n)\}|v = 240] \text{Prior}_{50}^{240}(n-1)}{P[\{e(n)\}]} 240 + \frac{P[\{e(n)\}|v = 490] \text{Prior}_{50}^{490}(n-1)}{P[\{e(n)\}]} 490$$

where n would be one and $P[\{e(n)\}] = P[\{e(n)\}|v = 240] \times \text{Prior}_{50}^{240}(n-1) + P[\{e(n)\}|v = 490] \times \text{Prior}_{50}^{490}(n-1)$. The initial prior probabilities are $\text{Prior}_{50}^{240}(0) = 0.45/0.65$ and $\text{Prior}_{50}^{490}(0) = 0.20/0.65$, respectively.

As traders received the hint “Not 50”, they know the value is either 240 or 490. Thus, it remains to calculate $P[\{e(1)\}|v]$ for $v \in \{240, 490\}$. This probability is equal to the probability that neither the current best bid nor the current best ask was traded given the value v .⁴² In this case the probability of the $\{76, 332\}$ bid-ask pair is equal to the following:

$$\begin{aligned} P[\{e(1)\}|v] &= P[\{\text{NT}(b_0) \cap \text{NT}(a_0)\} | v, b_0, a_0] \\ &= P[\text{NT}(b_0) | v, b_0, a_0] \times P[\text{NT}(a_0) | v, b_0, a_0, \text{NT}(b_0)] \times P[b_1 \in X_{b_1} \cap a_1 \in X_{a_1} | v, b_0, a_0] \\ &= P[b_1 > b_0 | v, b_0, a_0] \times P[a_1 < a_0 | v, b_0, a_0, b_1 > b_0] \times P[b_1 \in X_{b_1} \cap a_1 \in X_{a_1} | v, b_0, a_0] \quad (1) \end{aligned}$$

where the default initial bid, b_0 , is set to 50 and the default ask, a_0 , is 490. X_{b_1} and X_{a_1} represent the regions in which the new bid (76) and new ask (332) reside. We assume all bids and asks, and hence prices, reside in the interval $[50, 490]$. This interval is then divided into regions determined by the traders’ beliefs as depicted in the upper panel of Figure B2.1.

⁴² As previously noted in Appendix B1, this probability is equal to the product of three terms: (1) probability that the bid was not accepted given the value v , (2) probability that the ask was not accepted given the value v and the event that the bid was not traded and (3) probability that the new bid and ask lie in their respective regions.

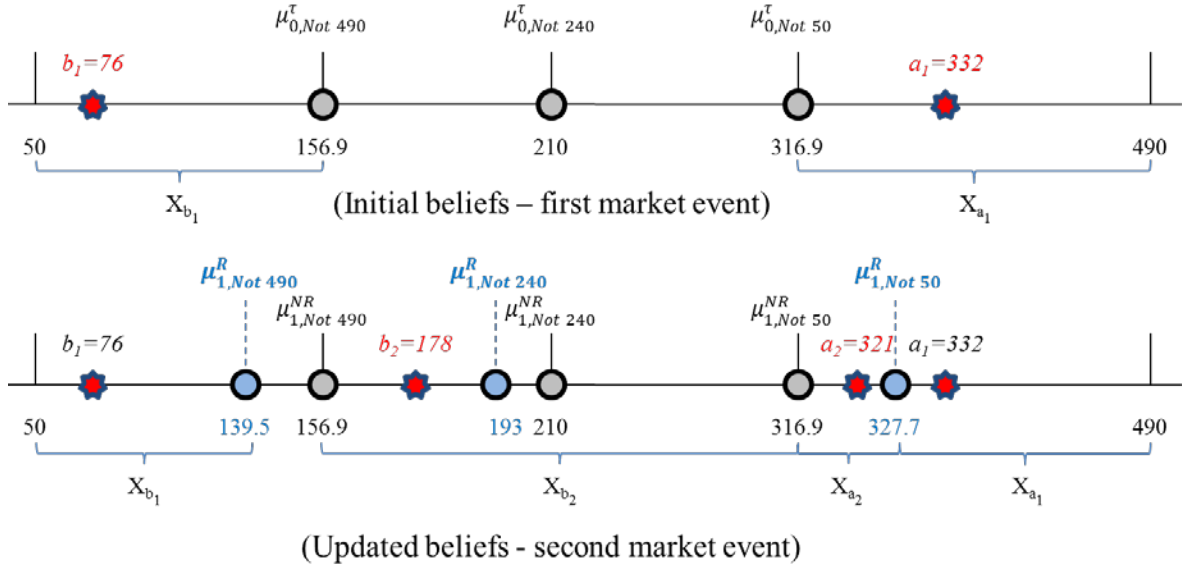


Figure B2.1. Representation of traders' beliefs before (upper panel) and after (lower panel) the initial bid-ask improvement event. The regions identified as X_{a_1} and X_{b_1} assume the true value of the asset is 240.

In words, equation (1) indicates that the probability of the bid-ask improvement given the true value of the asset is v is equal to the product of the following terms:

- [i] The probability that the new bid improves the old bid given the value of the asset is v , and given the old bid-ask pair.
- [ii] The probability that the new ask improves the old ask given the value, v , the old bid-ask pair (b_0, a_0) , and that the new bid improves the old bid.
- [iii] The probability that the new bid resides in the region $X_{b_1} = [50, 156.9]$ ($[50, 210]$) and the new ask is in $X_{a_1} = [316.9, 490]$ given the value of the asset is $v = 240$ (490), and given the old bid-ask pair.

We address each term in the product in order. Assume the true value is 240 (490). The term [i] is equivalent to one minus the probability that a bid in the region $[50, 156.9]$ ($[50, 210]$) was accepted/traded (i.e., one minus the probability of a trader selling in this region). *Reflective* traders assess this probability to be one, since all traders' beliefs are greater than 156.9 (210). The difference between regions when assuming the true value is 240 or 490 is due to the fact that

if the true value is assumed to be 240 (490), then no traders with the hint “Not 240” (“Not 490”) exist.

The term $[ii]$ is equivalent to one minus the ratio of the probability that a trader buys in the region $[316.9, 490]$ to the probability that a trader buys in the region $[50, 156.9]$ ($[50, 210]$ if the true value is assumed to be 490). From the upper panel in Figure B2.1 (first market event) it is clear that no trader is willing to buy in the former region, while all traders would seek to purchase the asset in the latter region. Thus, this term also equals one.

Finally, the term $[iii]$ consists of a product of two sums. The first sum is the probability of a trader buying in the region containing the new bid, which is equivalent to the probability that a trader would submit a bid in that region (i.e., her belief is greater than the upper bound of the region). The second sum, the probability of a trader selling in the region containing the new ask given the new bid, is calculated in the same manner. The *reflective* trader with hint “Not 50” knows the true value is either 240 or 490. Assuming it is 240, $[iii]$ is calculated as follows:

$$\left[\sum_{\substack{(\tau, h); \\ h \neq \text{"Not } v"}}} \frac{\gamma}{2} \left\{ \frac{\min(u_{x_{b_1}}, a_0) - \max(l_{x_{b_1}}, b_0)}{\min(\mu_{0,h}^\tau, a_0) - \max(50, b_0)} \right\} \right] \times \left[\sum_{\substack{(\tau, h); \\ h \neq \text{"Not } v"}}} \frac{\gamma}{2} \left\{ \frac{\min(u_{x_{a_1}}, a_0) - \max(l_{x_{a_1}}, b_1)}{\min(490, a_0) - \max(\mu_{0,h}^\tau, b_1)} \right\} \right]$$

If the trader *group*, being considered involves *reflective* traders, then the variable γ is set to α^τ , $\tau = R_P$ for *perceptive* or R_{NP} for *non-perceptive* traders. If the trader *group* being considered involves *non-reflective* traders, then the variable γ is set to $(1 - \alpha^\tau)$. When updating the belief of the *perceptive reflective* trader, α^τ is equal to $4/12$ ($\alpha^{R_P} = \alpha$); whereas α^τ equals $9/12$ ($\alpha^{R_{NP}}$) for the *non-perceptive reflective* trader in this example.

The assumption that *reflective* traders believe other *reflective* traders share their belief of the proportion of *reflective* traders in the market impacts a trader’s assessment of others’ beliefs. Assume the true value of the asset is 240 and consider the perspective of the *perceptive reflective* trader with the hint “Not 50.” To calculate $P[\{e(1)\} | v = 240]$, this trader must account for the beliefs of four *groups* of traders: *reflective* traders with the hint “Not 50”, *reflective* traders with the hint “Not 490”, *non-reflective* traders with the hint “Not 50” and *non-reflective* traders with the hint “Not 490”. Regions X_{b_1} and X_{a_1} are determined by the location of the new bid and new

ask and are set to [50,156.9] and [316.9,490], respectively (see upper panel in Figure B2.1, first market event). As we assume the true value is 240, note that no traders with the hint “Not 240” would exist.

Assuming the true value is 240, the probability that a *reflective* trader with the hint “Not 50” placed the bid, b_1 , in the region [50,156.9] is the proportion of such traders, $\frac{\alpha^\tau}{2}$, multiplied by $\left\{ \frac{\min(u_{X_{b_1},a_0}) - \max(l_{X_{b_1},b_0})}{\min(\mu_{0,Not\ 50}^R, a_0) - \max(50, b_0)} \right\} = \frac{156.9-50}{316.9-50}$. The probability that a *non-reflective* trader with the hint “Not 50” placed such a bid is given by $\frac{1-\alpha^\tau}{2}$ multiplied by $\frac{156.9-50}{316.9-50}$. Similarly, the probability that a *reflective* trader with the hint “Not 490” placed the bid in this region is given by $\frac{\alpha^\tau}{2} \frac{156.9-50}{156.9-50}$, while the probability that this resulted from the action of a *non-reflective* trader with hint “Not 490” is given by $\frac{1-\alpha^\tau}{2} \frac{156.9-50}{156.9-50}$. The probability that the bid was placed in this region given the true value of the asset is 240 is the sum of these terms.

Analogously, the probability that a trader with the hint “Not 50” placed the ask, a_1 , in the region [316.9,490] given the new bid is $\frac{1}{2} \left\{ \frac{\min(u_{X_{a_1},a_0}) - \max(l_{X_{a_1},b_1})}{\min(490, a_0) - \max(\mu_{0,Not\ 50}^\tau, b_1)} \right\} = \frac{1}{2} \frac{490-316.9}{490-316.9}$ where we have taken advantage of the fact that the *reflective* and *non-reflective* traders with the same hint initially hold the same belief so that $\frac{\alpha^\tau}{2} + \frac{(1-\alpha^\tau)}{2} = \frac{1}{2}$. The corresponding probability for traders with the hint “Not 490” is $\frac{1}{2} \frac{490-316.9}{490-156.9}$.

This allows one to calculate the term [iii] as:

$$\left[\frac{1}{2} \left(\frac{156.9-50}{316.9-50} \right) + \frac{1}{2} \left(\frac{156.9-50}{156.9-50} \right) \right] \times \left[\frac{1}{2} \left(\frac{490-316.9}{490-316.9} \right) + \frac{1}{2} \left(\frac{490-316.9}{490-156.9} \right) \right] = 0.53.$$

A similar calculation yields 0.65 when the *perceptive reflective* trader with the hint “Not 50” assumes the true value is 490. Thus, this trader updates her belief to:

$$\begin{aligned} \mu_{1,Not\ 50}^R &= \frac{(1)(1)(0.53)(0.45/0.65)}{(0.53)(0.45/0.65) + (0.65)(0.2/0.65)} 240 + \frac{(1)(1)(0.65)(0.2/0.65)}{(0.53)(0.45/0.65) + (0.65)(0.2/0.65)} 490 \\ &= 327.7 \end{aligned}$$

As the *reflective* and *non-reflective* traders with the same hints share the same initial belief, this first update is independent of the traders' belief of the proportion of *reflective* traders in the market (i.e., there is no difference between the updates of *perceptive reflective* and *non-perceptive reflective* traders). Following the same process outlined above, *reflective* traders with the hint “Not 490” (“Not 240”) update their beliefs from 156.9 (210) to 139.5 (193). Recall that the *non-reflective* traders do not update their beliefs (see lower panel in Figure B2.1, second market event).

Second market event: bid-ask spread {178, 321}

The second randomly selected trader will not accept the bid or ask, as all traders' beliefs reside within the bid-ask spread {76, 332} (see upper panel in Figure B2.1, first market event). Suppose the selected trader improves the spread to {178, 321}.⁴³ The belief updates will now depend upon the *reflective* traders' belief of the proportion of *reflective* traders in the market. Indeed, consider the *perceptive reflective* trader with the hint “Not 50”. We again must calculate [i], [ii] and [iii] defined above. Assume the true value is 240 so that the regions of interest are as depicted in the lower panel of Figure B2.1. The probability that a bid in the region [50, 139.5] would be accepted is zero as all traders' beliefs are greater than 139.5. Thus, the term [i] is equal to 1. The probability that an ask in the region [327.7, 490] would be accepted is also zero as all traders' beliefs are less than or equal to 327.7. Thus, term [ii] is equal to 1. It remains to calculate [iii].

The probability that a *reflective* trader with the hint “Not 50” placed the bid, $b_2 = 178$, in the region [156.9, 316.9] is the proportion of such traders, $\frac{\alpha^\tau}{2}$, multiplied by $\left\{ \frac{\min(u_{x_{b_2}}, a_1) - \max(l_{x_{b_2}}, b_1)}{\min(\mu_{1, \text{Not } 50}^R, a_1) - \max(50, b_1)} \right\} = \frac{316.9 - 156.9}{327.7 - 76}$. The probability that a *non-reflective* trader with the hint “Not 50” placed such a bid is given by $\frac{1 - \alpha^\tau}{2}$ multiplied by $\frac{316.9 - 156.9}{316.9 - 76}$. The probability that a *reflective* trader with the hint “Not 490” placed the bid in this region is zero since her belief of

⁴³ Note that this second trader cannot be a *reflective* trader with the hint “Not 50,” as the updated ask of 321 is less than the belief of the *reflective* traders with the hint “Not 50”.

the true value is less than the upper bound of the region. The corresponding probability for the *non-reflective* trader with hint “Not 490” is also zero for the same reason. The probability that the bid was placed in this region given the true value of 240 is the sum of these terms.

Analogously, the probability that a *reflective* trader with the hint “Not 50” placed the ask, $a_2 = 321$, in the region $[316.9, 327.7]$ given the bid of 178 is zero as her belief is greater than the lower bound of the region. The probability that a *non-reflective* trader with the hint “Not 50” placed such an ask is given by $\frac{1-\alpha^\tau}{2} \left\{ \frac{\min(u_{X_{a_2}, a_1}) - \max(l_{X_{a_2}, b_2})}{\min(490, a_1) - \max(\mu_{1, \text{Not } 50}^{NR}, b_2)} \right\} = \frac{1-\alpha^\tau}{2} \frac{327.7-316.9}{332-316.9}$. Similarly, the probability that a *reflective* trader with the hint “Not 490” placed the ask in this region is $\frac{\alpha^\tau}{2} \frac{327.7-316.9}{332-178}$, while the probability that a *non-reflective* trader with the hint “Not 490” submitted this ask is $\frac{1-\alpha^\tau}{2} \frac{327.7-316.9}{332-178}$. Thus, term [iii] is calculated as:

$$\left[\frac{\alpha^\tau}{2} \left(\frac{316.9 - 156.9}{327.7 - 76} \right) + \frac{1-\alpha^\tau}{2} \left(\frac{316.9 - 156.9}{316.9 - 76} \right) + \frac{\alpha^\tau}{2} (0) + \frac{1-\alpha^\tau}{2} (0) \right] \\ \times \left[\frac{\alpha^\tau}{2} (0) + \frac{1-\alpha^\tau}{2} \left(\frac{327.7 - 316.9}{332 - 316.9} \right) + \frac{1}{2} \left(\frac{327.7 - 316.9}{332 - 178} \right) \right] = 0.196$$

where $\alpha^\tau = \alpha^{RP} = 4/12$. A similar calculation yields 0.0896 when the *perceptive reflective* trader with the hint “Not 50” assumes the true value is 490.⁴⁴ Thus, this trader updates her belief to:

$$\mu_{2, \text{Not } 50}^{RP} = \frac{(1)(1)(0.0896)(0.65)}{(0.0896)(0.65) + (0.196)(0.35)} 240 + \frac{(1)(1)(0.196)(0.35)}{(0.0896)(0.65) + (0.196)(0.35)} 490 = 375.5$$

where 0.65 (0.35) is the prior probability that the value is 240 (490). Note that the trader’s belief of the actual proportion of *reflective* traders in the market matters in this case. Indeed, given $\alpha^{RNP} = 9/12$ for the *non-perceptive reflective* trader, her updated belief of the asset value is 379.4. Note that both traders updated their beliefs in the “wrong” direction. Similarly, the *perceptive reflective* trader with the hint “Not 490” updated her belief to 124.5, while the *non-*

⁴⁴ If the true value is assumed to be 490, then traders with the hint “Not 490” would not exist. Thus, the region in which the bid of 178 resides would be $[50, 193]$ instead of $[156.9, 316.9]$. Similarly, in the calculation of term [i] the region in which the bid of 76 resides would be $[50, 193]$ instead of $[50, 139.5]$.

perceptive reflective trader with the hint “Not 490” revised her belief to 91.6. The traders’ updated beliefs are reflected in Figure B2.2.

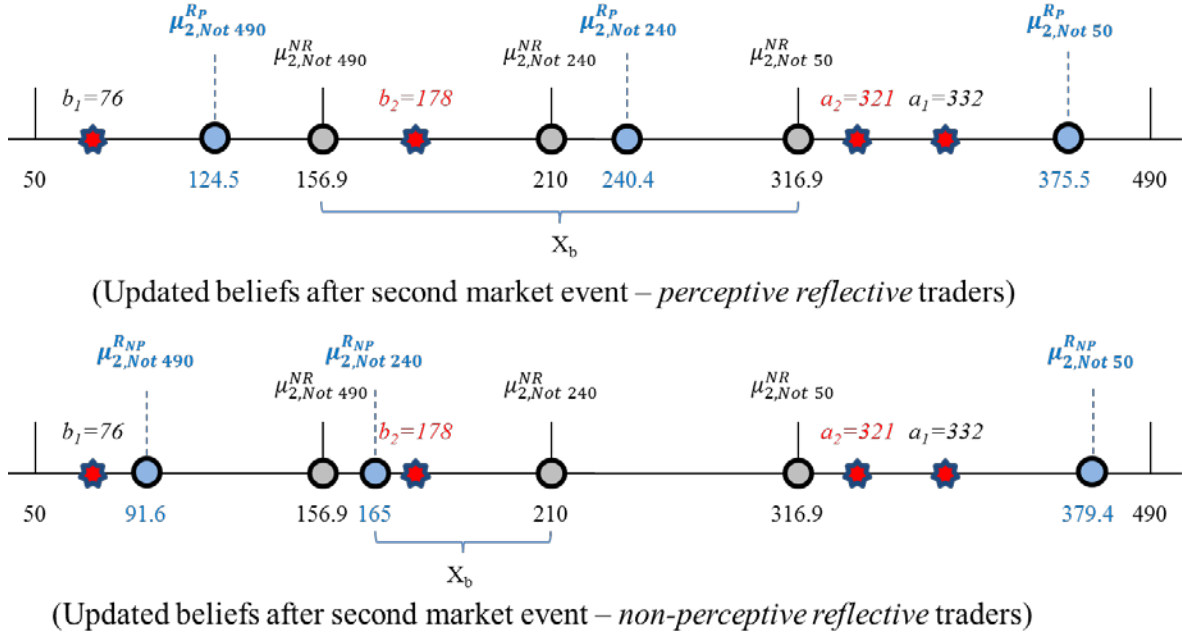


Figure B2.2. Representation of *perceptive reflective* (upper panel) and *non-perceptive reflective* (lower panel) traders’ beliefs after the second bid-ask improvement event. The regions identified as X_b assume the true value of the asset is 240.

Third market event: sale at 178

To conclude this descriptive example, suppose the bid of 178 is accepted by the third randomly selected trader. Consider the perspective of the *perceptive reflective* trader with the hint “Not 50.” She updates her belief as follows:

$$P[\{e(3)\}|v] = P[\{T(b_2)\}|v] = P[a \text{ traders sells in } X_{b_2}|v] = \rho_{sell}^{X_{b_2}}$$

To calculate the probability of selling in region X_{b_2} , set $\rho_{sell}^{X_{b_2}} = 0$ and consider the belief of each trader *group*. Suppose the true value of the asset is 240. If the belief of *reflective* (*non-reflective*) traders with the hint “Not 50” is less than or equal to the lower bound of the region within which the current best bid of 178 resides, i.e. $[156.9, 316.9]$, then add $\frac{\alpha^{RP}}{2} = \frac{4/12}{2} \left(\frac{1-\alpha^{RP}}{2} = \frac{8/12}{2} \right)$ to $\rho_{sell}^{X_{b_2}}$.

We proceed similarly, for traders with the hint “Not 490.”

Thus,

$$P[\{e(3)\}|240] = \rho_{sell}^{x_{b_2}} = \frac{1}{2}$$

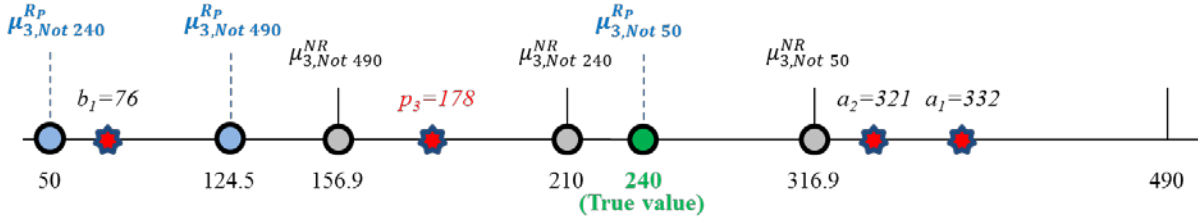
as both *reflective* and *non-reflective* traders with the hint “Not 490” would be willing to sell in this region while traders with the hint “Not 50” would not. Analogously, if we assume the true value of the asset is 490, then:

$$P[\{e(3)\}|490] = \rho_{sell}^{x_{b_2}} = 0$$

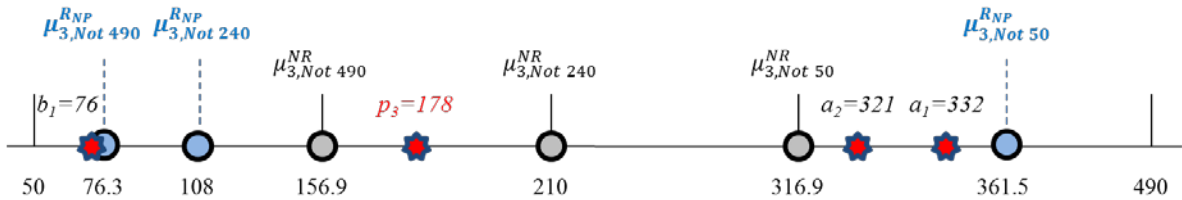
for the region [50,210] (see the upper panel in Figure B2.2). Thus, the *perceptive reflective* trader with the hint “Not 50” learns the true value is 240. However, the *non-perceptive reflective* trader with hint “Not 50” fails to learn the true value of the asset updating her belief to:

$$\mu_{3,50}^{RNP} = \frac{(1/2)(0.443)}{(1/2)(0.443) + (3/8)(0.557)} 240 + \frac{(3/8)(0.557)}{(1/2)(0.443) + (3/8)(0.557)} 490 = 361.5.$$

The *perceptive reflective* trader with the hint “Not 490” updates her belief to 124.5, and her *non-perceptive reflective* counterpart updates her belief to 76.3. Thus, this example demonstrates the importance of a trader’s ability to accurately perceive the proportion of *reflective* traders in the market (compare the upper and lower panels in Figure B2.3).



(Updated beliefs after third market event – *perceptive reflective* traders)



(Updated beliefs after third market event – *non-perceptive reflective* traders)

Figure B2.3. Representation of *perceptive reflective* (upper panel) and *non-perceptive reflective* (lower panel) traders’ beliefs after the first transaction.

Note that the new current best bid is 76 (this is the case because, as in the experiment, we do not clear the order book after a transaction), while the current best ask remains 321. The simulation would continue with the random selection of a fourth trader.

Appendix B3. Simulations and hypotheses⁴⁵

Table B3.1. Average results of 25,000 simulations for each $(v, \alpha, \alpha^{RNP})$ for earnings and proportion of consistent orders across behavioral types and hints (Hypotheses 1 and 2)

	(A) <i>Earnings (francs)</i>			(B) <i>Proportion of consistent orders</i>		
	All hints	Hint “Not 240”	Hint “Not 50” “Not 490”	All hints	Hint “Not 240”	Hint “Not 50” “Not 490”
<i>Perceptive reflective</i>	2,269.25	2,311.75	2,247.97	0.814	0.832	0.812
<i>Non-perceptive reflective</i>	2,265.07	2,307.20	2,244.00	0.795	0.821	0.790
<i>Non-reflective</i>	2,210.32	2,172.78	2,229.65	0.442	0.259	0.544

The number in each cell corresponding to a *reflective* trader (first two rows) reflects the average value across $3 \times 11 \times 13 \times 25,000$ simulations. As non-reflective traders do not exist when $\alpha = 12/12$, the last row reports average values across $3 \times 10 \times 13 \times 25,000$ simulations

⁴⁵ The simulation program was coded in Matlab and is available upon request from the authors.

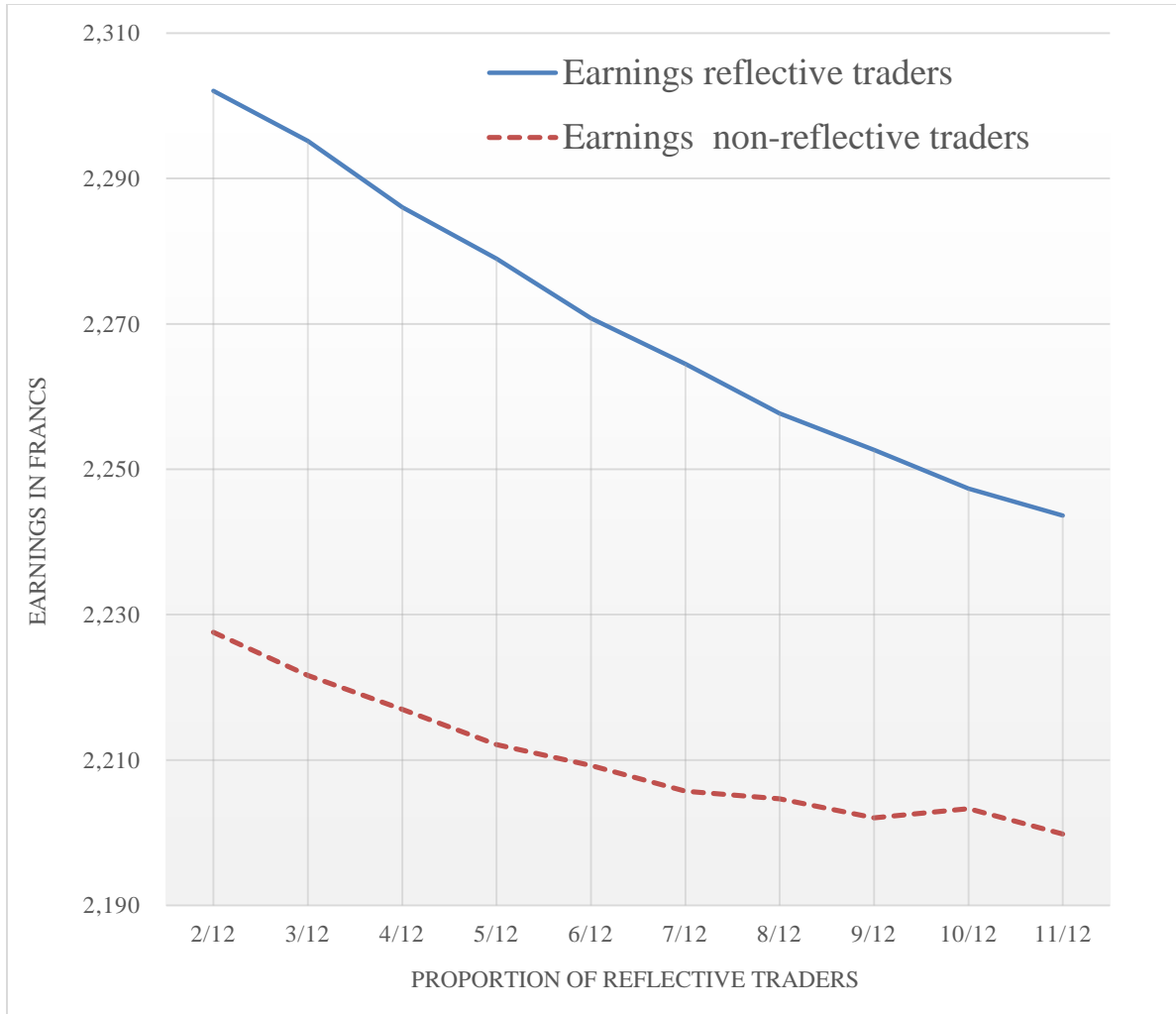


Figure B3.1. Earnings (in francs) of *reflective* (upper curve) and *non-reflective* traders (dotted lower curve) as a function of the proportion of *reflective* traders in the market (Hypothesis 1i) ranging from $\frac{2}{12}$ to $\frac{11}{12}$ to make sure all three behavioral types are present in the market.

Table B3.2. Average results of 25,000 simulations for each $(v, \alpha, \alpha^{RNP})$ combination for shareholdings at the end of the market across behavioral types and asset value (Hypothesis 3)

Behavioral type	$v = 50$	$v = 490$
<i>Perceptive reflective</i>	3.65	4.32
<i>Non-perceptive reflective</i>	3.69	4.28
<i>Non-reflective</i>	4.45	3.59

Appendix C. Additional analyses

Appendix C1. Winner's curse traps

We define winner's curse traps following Biais et al. (2005) by considering situations in which the true value of the asset is not 240, and the participant's signal does not rule out 240. In this case winner's curse traps manifest themselves when prices are close to 240 as agents are likely to wrongly infer that the true value of the asset is 240. Because the authors used an opening call auction in a large number of their experimental sessions they considered that a winner's curse trap existed when the opening call auction price was strictly larger than 220 and lower than or equal to 260. We adapt this working definition to our continuous double auction design by defining the winner's curse trap situation as one in which average prices were between 220 and 260 in the first minute of each market period. We report our findings in Table C1.1 below. We report our findings for this definition of winner's curse traps as well as for a slightly broader definition that includes first-minute prices between 200 and 280. We do so because, using the original definition only 31.0% of first-minute prices fall into the range [220-260] compared to 44.6% for the range [200-280]. However, the qualitative nature of the results is unaffected by the price range choice. Our findings are in line with Biais et al. (2005).

Table C1.1. Trader earnings as a function of individual characteristics and market variables

<i>Earnings (in \$)</i>				
Sample:	Winner's curse trap [220-260]	No Winner's curse trap [220-260]	Winner's curse trap [200-280]	No Winner's curse trap [200-280]
Intercept	1,173.308**** (99.318)	1,041.564**** (50.228)	1,183.284**** (67.167)	1,057.102**** (51.889)
<i>Hypothesized predictors</i>				
ToM Top 25% Dummy	277.870** (144.928)	82.107*** (27.862)	289.02*** (110.436)	63.876** (30.545)
<i>Controls</i>				
<i>Individual characteristics</i>				
CRT score	101.373* (53.180)	39.613** (17.693)	70.766* (40.278)	42.347** (18.395)

Raven score	43.539 (62.690)	45.996** (22.098)	28.158 (49.622)	46.310** (22.292)
Financial literacy score	-60.494 (42.022)	15.916 (13.951)	-30.767 (43.348)	26.364 (23.227)
Male Dummy	1.997 (127.84)	78.235** (40.315)	45.342 (99.833)	68.496* (42.577)
<i>Market characteristics</i>				
Loan Dummy	-519.461**** (74.081)	-875.497**** (39.364)	-882.271**** (218.870)	-863.726**** (34.076)
Hint “Not 50” Dummy	— ⁴⁶	184.510**** (45.555)	—	176.982**** (45.188)
Hint “Not 490” Dummy	—	229.552**** (15.350)	—	226.128**** (15.653)
Asset value	4.112**** (0.214)	3.766**** (0.148)	4.072**** (0.218)	3.731**** (0.162)
Observations	n = 227	n = 2,612	n = 326	n = 2,513
Prob > χ^2	0.000	0.000	0.000	0.000
R ²	0.759	0.638	0.772	0.626

The coefficient associated to theory of mind (ToM Top 25% Dummy) is greater in the presence of winner’s curse traps than in their absence. To assess the statistical significance of this difference we conducted two additional regressions similar to the ones in Table C1.1 in which we included the whole sample and in which we added a winner’s curse dummy and an interaction dummy between theory of mind and the winner’s curse trap (ToM Top 25% Dummy \times Winner’s curse Dummy).⁴⁷ We found this interaction term to be statistically significant for the case in which we define the winner’s curse trap using the interval [200,280] (p-value = 0.011) and to be close to significance when using the interval [220,260] (p-value = 0.136). Statistical significance is most likely not achieved in the latter case because we only have 227 observations out of 2,839 falling into the winner’s curse trap category.

Appendix C2. Imprecise (“Not 240”) vs. precise (“Not 50” or “Not 490”) hints

⁴⁶ We do not introduce dummies for hints in the case of winner’s curse traps regressions as it would generate collinearity issues.

⁴⁷ These regression results are available upon request from the authors.

Table C2.1. Earnings and proportion of consistent orders as a function of individual characteristics and market variables when traders received (or did not receive) imprecise hints

Sample: <i>Perceptive reflective and non-reflective traders</i>	<i>Earnings</i>		<i>Proportion of consistent orders</i>	
	<i>Imprecise hints [1]</i>	<i>Precise hints [2]</i>	<i>Imprecise hints [3]</i>	<i>Precise hints [4]</i>
Intercept	970.812**** (65.453)	1,135.856**** (65.453)	0.618**** (0.074)	0.618**** (0.074)
<i>Hypothesized predictors</i>				
CRT & ToM Top 25% Dummy ⁴⁸	236.289*** (91.596)	139.194* (81.944)	0.113** (0.045)	0.048 (0.042)
<i>Controls</i>				
<i>Individual characteristics</i>				
Raven score	44.815 (47.065)	37.043 (32.165)	0.030 (0.020)	0.020 (0.013)
Financial literacy score	66.350 (52.815)	18.628 (26.972)	0.032 (0.024)	0.021 (0.014)
Male Dummy	64.922 (62.908)	190.427 (57.587)	0.003 (0.0666)	-0.014 (0.025)
<i>Market characteristics</i>				
Loan Dummy	-594.984**** (60.613)	-708.206**** (43.205)	-0.227** (0.103)	-0.010 (0.039) (0.103)
High stakes Dummy	—	—	0.021 (0.044)	-0.051*** (0.017)
Hint “Not 50” Dummy	—	—	—	-0.010 (0.055)
Asset value	3.824**** (0.295)	3.993**** (0.267)	-0.001 (0.001)	-0.001 (0.001)
Consistent with hint	—	—	0.247*** (0.081)	0.427**** (0.055)
Number of transactions	—	—	-0.005* (0.003)	-0.003 (0.002)
Market number	—	—	-0.004* (0.003)	-0.001 (0.002)

⁴⁸ This dummy variable takes value one if a trader scores in the top 25% of all participants in terms of both CRT and ToM scores.

			(0.002)	(0.001)
Observations ⁴⁹	n = 306	n = 816	n = 286	n = 781
Prob > χ^2	0.000	0.000	0.000	0.000
R ²	0.689	0.639	0.132	0.259

Appendix C3. Proportion of consistent orders

In our regression analysis of the proportion of consistent orders we follow a similar approach as for earnings. We add several control variables which are relevant in the case of the proportion of consistent orders.⁵⁰ We control for the proportion of orders which are consistent with a trader's own private information so that any significant effect of trader skills on the proportion of consistent trades cannot only be due to more consistent use of private information. We define an order to be consistent with one's private information if it is a bid or a purchase (an ask or a sale) that is lower (higher) than or equal to the expected value of the asset given one's own hint.

We also control for the number of transactions as well as the market number (17 markets are conducted each session) in each regression. In addition, we assess whether high stakes sessions are characterized by a higher proportion of orders consistent with the true value of the asset using a High stakes Dummy variable which takes value one if a trader was involved in a high stakes session. However, high stakes sessions appear to lead to a lower consistency of orders.⁵¹

Table C3.1. Proportion of consistent orders as a function of individual characteristics and market variables

	All	Top & Bottom 25% CRT
Sample:	[1]	[2]

⁴⁹ The difference in observations between columns 1 and 3 (analogously 2 and 4) is due to the fact that some traders might not trade during a market period. Thus, these traders are excluded from the proportion of consistent orders regressions.

⁵⁰ Keeping exactly the same regressors as in the earnings analysis does not change the qualitative nature of the results.

⁵¹ None of these additional controls were relevant for the analysis of earnings as the sum of individual earnings (once we divide high stakes sessions earnings by two) is the same across periods (with the same asset value), sessions and treatments.

Intercept	0.453**** (0.026)	0.420**** (0.034)
<i>Hypothesized predictors</i>		
CRT score	0.029**** (0.006)	—
CRT Top 25% Dummy	—	0.055**** (0.017)
<i>Controls</i>		
<i>Individual characteristics</i>		
Raven score	0.012* (0.007)	0.012* (0.007)
Financial literacy score	0.010 (0.007)	0.019* (0.010)
Male Dummy	-0.004 (0.011)	0.006 (0.012)
<i>Market characteristics</i>		
Loan Dummy	0.001 (0.010)	0.008 (0.014)
High stakes Dummy	-0.025** (0.011)	-0.044**** (0.014)
Hint “Not 50” Dummy	0.150**** (0.029)	0.169**** (0.031)
Hint “Not 490” Dummy	0.180**** (0.025)	0.175**** (0.028)
Asset value	-0.001** (0.001)	-0.001** (0.001)
Consistent with hint	0.197**** (0.018)	0.199**** (0.032)
Number of transactions	-0.006**** (0.001)	-0.006**** (0.001)
Market number	-0.001 (0.001)	-0.001 (0.001)
Observations	n = 3,267	n = 2,086
Prob > χ^2	0.000	0.000
R ²	0.158	0.161

Table C3.2. Proportion of consistent orders as a function of individual characteristics (including theory of mind) and market variables

<i>Proportion of consistent orders</i>					
Sample:	All	All	Top & Bottom 25% ToM	Top 25% CRT	Bottom 25% CRT
	[1]	[2]	[3]	[4]	[5]
Intercept	0.450**** (0.030)	0.447**** (0.030)	0.473**** (0.043)	0.387**** (0.064)	0.352**** (0.048)
<i>Hypothesized predictors</i>					
ToM score	0.010 (0.009)	0.014** (0.006)	—	0.042*** (0.016)	-0.007 (0.009)
ToM Top 25% Dummy	—	—	0.047** (0.020)	—	—
ToM score × CRT score	—	0.018**** (0.005)	—	—	—
CRT score	0.032**** (0.008)	0.032**** (0.007)	0.030**** (0.008)	0.083*** (0.024)	-0.027 (0.034)
<i>Controls</i>					
<i>Individual characteristics</i>					
Raven score	0.016** (0.007)	0.018*** (0.007)	0.008 (0.007)	0.012 (0.014)	0.019** (0.009)
Financial literacy score	0.014** (0.007)	0.014** (0.007)	0.013 (0.009)	0.009 (0.012)	0.041**** (0.010)
Male Dummy	-0.002 (0.010)	-0.005 (0.012)	0.002 (0.016)	-0.003 (0.027)	-0.015 (0.027)
<i>Market characteristics</i>					
Loan Dummy	0.014 (0.010)	0.015* (0.008)	0.024 (0.022)	0.043** (0.019)	0.084** (0.040)
High stakes Dummy	-0.015 (0.010)	-0.018* (0.011)	-0.040*** (0.012)	-0.022*** (0.026)	-0.025 (0.027)
Hint “Not 50” Dummy	0.149**** (0.028)	0.147**** (0.028)	0.183**** (0.040)	0.167**** (0.037)	0.147**** (0.037)
Hint “Not 490” Dummy	0.180**** (0.028)	0.180**** (0.028)	0.209**** (0.037)	0.175**** (0.036)	0.165**** (0.039)
Asset value	-0.0002** (0.0001)	-0.0002** (0.0001)	-0.0003** (0.0001)	-0.0001 (0.0001)	-0.0002* (0.0001)
Consistent with hint	0.197**** (0.022)	0.203**** (0.023)	0.155**** (0.025)	0.176**** (0.051)	0.249**** (0.052)
Number of transactions	-0.006**** (0.001)	-0.006**** (0.001)	-0.006**** (0.001)	-0.006**** (0.001)	-0.005**** (0.002)
Market number	-0.002**	-0.002**	-0.002	-0.001	-0.001

	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	n = 2,698	n = 2,698	n = 1,451	n = 1,000	n = 767
Prob > χ^2	0.000	0.000	0.000	0.000	0.000
R ²	0.161	0.164	0.194	0.157	0.179

Appendix C4. Shareholdings

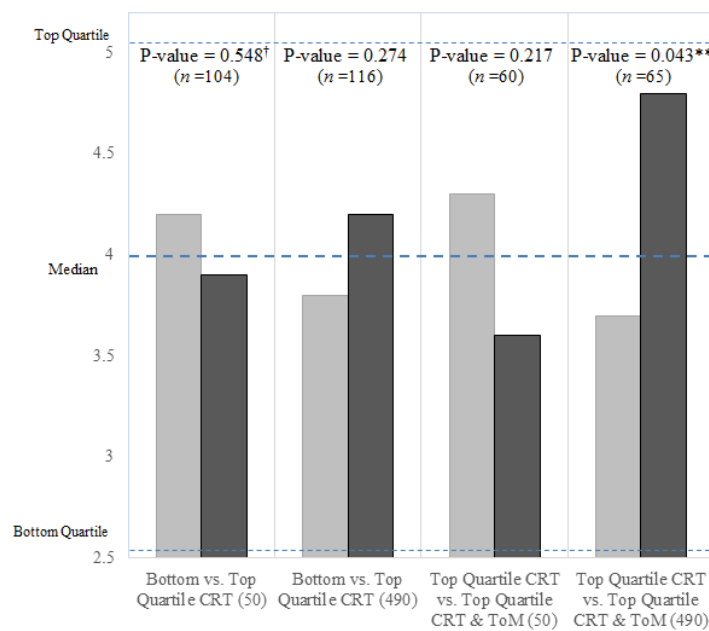


Figure C4.1. Average final shareholdings for participants in the bottom and top quartiles for CRT and ToM depending on the true value of the asset (50 or 490). We also represent the bottom quartile, median and top quartile for the whole distribution of participants' shareholdings (see dashed horizontal lines).

†P-values based on WRS comparing shareholdings of bottom and top quartiles traders in terms of CRT, ToM, and both CRT and ToM. *p-value<0.10, **p-value<0.05, ***p-value<0.01 and ****p-value<0.001

Table C4.1. Net holdings consistency as a function of individual characteristics (including theory of mind) and market variables

<i>Net holdings consistency</i>				
Sample:	All	All	Top 25% CRT	Bottom 25% CRT

	[1]	[2]	[3]	[4]
Intercept	-0.537*** (0.194)	-0.786*** (0.269)	-2.470** (1.166)	-0.550 (0.440)
<i>Hypothesized predictors</i>				
CRT score	0.264**** (0.070)	0.259*** (0.082)	1.105** (0.462)	-0.025 (0.387)
CRT Top 25% Dummy	—	—	—	—
ToM score	—	0.080* (0.046)	0.527*** (0.201)	-0.148 (0.126)
ToM Top 25% Dummy	—	—	—	—
ToM score × CRT score	—	0.254**** (0.074)	—	—
<i>Controls</i>				
<i>Individual characteristics</i>				
Raven score	0.067 (0.069)	0.112 (0.094)	0.362 (0.262)	0.054 (0.115)
Financial literacy score	0.119 (0.091)	0.115 (0.104)	0.051 (0.160)	0.285** (0.126)
Male Dummy	-0.148 (0.143)	-0.124 (0.190)	-0.096 (0.266)	-0.165 (0.369)
<i>Market characteristics</i> ⁵²				
Loan Dummy	0.240** (0.107)	0.323** (0.133)	0.147 (0.185)	1.122** (0.544)
High stakes Dummy	0.118* (0.063)	0.208*** (0.071)	0.271 (0.366)	0.060 (0.181)
Hint “Not 50” Dummy	1.867**** (0.440)	1.909**** (0.467)	1.931** (0.889)	1.100** (0.493)
Hint “Not 490” Dummy	1.722**** (0.404)	1.907**** (0.455)	2.049*** (0.696)	1.044**** (0.282)
Asset value	-0.001 (0.001)	-0.001 (0.001)	0.002 (0.002)	-0.001 (0.001)
Number of transactions	-0.049** (0.025)	-0.048* (0.028)	-0.050**** (0.014)	-0.012 (0.025)
Observations	n = 2,040	n = 1,670	n = 600	n = 480
Prob > χ^2	0.000	0.000	0.000	0.000
R ²	0.056	0.061	0.065	0.058

⁵² We do not control for market number which turns out not to be statistically significant in any of the specifications used in the table.

Appendix D. Further inquiries into the model hypotheses

Appendix D1. CRT and Bayesian updating

To confirm that the positive effect of CRT on earnings is driven by the fact that high-CRT traders are immune to Bayesian updating mistakes, we conduct regressions in which we include Bayesian updating scores (following Charness and Levin, 2009, see Appendix A for the details of the test) as an independent variable. We collected this variable in five sessions which corresponds to a total of 60 traders. Because Bayesian updating scores are positively and significantly correlated with CRT scores ($r = 0.313$, $p\text{-value} = 0.014$) we orthogonalize the two variables to perform our regression analyses. In particular, we isolate the effect of Bayesian updating performance from CRT score performance by orthogonalizing CRT scores with respect to Bayesian updating scores. We find that Bayesian updating test scores explain earnings, the proportions of consistent orders as well as net holdings consistency whereas orthogonalized CRT scores do not (see Table D1.1). It follows that, in line with our theoretical hypotheses, the explanatory power of CRT lies in its capacity to explain individuals' Bayesian updating performance.

Table D1.1. Earnings as a function of extended individual characteristics and market variables

Sample:	Earnings	Orders Consistency	Net holdings consistency
Intercept	991.665**** (33.862)	0.467**** (0.042)	-0.528 (0.396)
<i>Hypothesized predictors</i>			
Bayesian updating score	38.36** (18.517)	0.019**** (0.003)	0.098** (0.047)
CRT score	34.082 (45.081)	0.031 (0.019)	0.208 (0.173)
<i>Controls</i>			
<i>Individual characteristics</i>			
Raven score	33.308**** (8.777)	0.017 (0.011)	0.026 (0.140)
Financial literacy score	16.537 (27.332)	0.026**** (0.007)	0.264** (0.113)
Male	143.540*** (48.421)	0.016 (0.027)	0.338 (0.366)

<i>Market characteristics</i>			
Hint “Not 50”	221.131**** (85.386)	0.113*** (0.042)	1.624** (0.665)
Hint “Not 490”	224.110**** (33.728)	0.104** (0.053)	1.298*** (0.387)
Asset value	4.032**** (0.093)	-0.001 (0.001)	-0.001 (0.001)
Consistent with hint	—	0.191**** (0.035)	—
Number of transactions	—	-0.005** (0.002)	-0.032 (0.024)
Observations	1,020	988	600
Prob > χ^2	0.000	0.000	0.000
R ²	0.658	0.116	0.077

Appendix D2. Theory of mind and CRT rank beliefs

To further establish the soundness of our model, we aim to connect the positive effect of theory of mind skills to one’s capacity to assess other traders’ levels of cognitive reflection. To that end, we first compute a variable capturing the accuracy of traders’ beliefs regarding their own ranking in terms of CRT scores. Due to the follow-up survey, we have individual information regarding a subject’s *perception* of their CRT rank in the population of 1,642 students in the lab subject pool. Because we know all CRT scores of the pool of subjects in our database, we can also compute the absolute difference between one’s *perceived* CRT rank and one’s *actual* rank. We refer to this measure as the *CRT beliefs mismatch*. In line with our model, this measure is significantly and negatively correlated with ToM scores ($r = -0.28$, $p\text{-value} < 0.001$). It is worth noting that this measure also correlates positively and significantly with CRT-based overconfidence ($r = 0.23$, $p\text{-value} = 0.003$) as most subjects (63.2%) who have positive *CRT beliefs mismatch* are also overconfident, *perceiving* their rank to be better than it actually is.⁵³

For each session, we can also calculate the absolute difference between one’s *perceived* CRT rank in the lab subject pool and their *actual* CRT rank in a given session. This captures the

⁵³ The CRT-based overconfidence measure is calculated, in the spirit of Svenson (1981), by taking the difference between one’s *actual* CRT rank and one’s *perceived* CRT rank using as reference the pool of subjects eligible to participate in experiments at the lab where the study was conducted ($n = 1,642$ at the time of the study).

mismatch between a trader's belief regarding their *perceived* CRT rank and their actual rank in the market. We refer to this variable as *CRT beliefs market mismatch*. In line with our model, *CRT beliefs market mismatch* affects earnings negatively (see Table D2.1, column [1]). We also show that our CRT-based overconfidence measure does not affect earnings once it is orthogonalized with respect to *CRT beliefs market mismatch* (see Table D2.1, column [4]).

Interestingly, the negative effect of the variable *CRT beliefs market mismatch* is limited to the case of *non-perceptive* traders (see Table D2.1). This finding is consistent with the fact that *perceptive* traders may be more able to adjust their beliefs to current market conditions than *non-perceptive* traders. Thus, *perceptive* traders may not suffer from initially misleading CRT ranking beliefs.

In the case in which the variable *CRT beliefs market mismatch* is equal to zero, traders' beliefs regarding their CRT rank correspond to their *actual* CRT rank in the market. This resembles the situation in our model in which $\alpha^{RNP} = \alpha$.⁵⁴ In the case in which traders hold accurate beliefs about their *actual* CRT rank in the market we expect theory of mind to be less relevant. In line with this conjecture, we find that the positive effect of ToM scores only arises when traders have inaccurate beliefs about their *actual* CRT rank (see columns [2] and [3] Table D2.2).

In our model, high theory-of-mind skills traders, which we refer to as *perceptive*, have a more precise assessment of the proportion of *reflective* traders in the market than low theory-of-mind skills traders. To obtain a direct measure of the ability of traders to assess the composition of behavioral types in the market, we ask them, at the end of the experiment, how they would rank themselves on the CRT compared to the other traders in the market (see Appendix A). We then compute the distance between their *perceived* rank and their *actual* rank as a way to assess the accuracy of their beliefs regarding other traders' CRT scores. We consider that, the larger the distance the less accurate a trader's belief is regarding the distribution of CRT scores in the market.⁵⁵ In line with the previous argument, this measure correlates positively with theory of mind skills ($r = 0.174$). This positive correlation does not reach statistical significance (p-value

⁵⁴ We do not elicit traders' beliefs about the distribution of behavioral types so we cannot ensure that $\alpha^{RNP} = \alpha$.

⁵⁵ We do not elicit the whole distribution of beliefs regarding CRT scores, however.

= 0.236), however, as we collected the end-of-experiment CRT rank elicitation measure for only a subset of 48 subjects which limits the power of our statistical analysis.

Table D2.1. Earnings as a function of individual characteristics and CRT beliefs market mismatch

Sample:	All [1]	Bottom 25% ToM [2]	Top 25% ToM [3]	All [4]
Intercept	1,054.390**** (45.469)	1,070.224**** (138.890)	1,138.749**** (144.999)	1,050.912**** (46.237)
<i>Hypothesized predictors</i>				
CRT beliefs market mismatch	-27.341** (11.071)	-42.821** (17.303)	-10.450 (36.343)	-27.421*** (10.031)
CRT-based overconfidence	—	—	—	-25.845 (16.853)
ToM score	33.353*** (12.471)	33.007 (21.542)	67.049 (107.140)	34.209** (13.487)
CRT score	41.045** (18.959)	-19.417 (31.814)	55.762 (35.008)	39.955** (16.620)
<i>Controls</i>				
<i>Individual characteristics</i>				
Raven score	44.655** (18.092)	64.127*** (23.102)	-16.708 (34.415)	44.661** (18.267)
Financial literacy score	8.506 (14.687)	-26.824 (26.084)	15.633 (21.966)	8.053 (14.948)
Male	75.089** (36.330)	207.752**** (59.325)	59.715 (57.314)	82.979** (36.387)
<i>Market characteristics</i>				
Loan Dummy	-865.336**** (32.789)	-715.721**** (32.823)	-870.213**** (60.044)	-862.288**** (29.142)
Hint “Not 50”	198.278**** (45.551)	233.428** (91.735)	181.879** (77.035)	197.379**** (45.000)
Hint “Not 490”	234.096**** (16.264)	261.529**** (54.072)	174.684*** (65.085)	234.554**** (16.130)
Asset value	3.808**** (0.135)	3.293**** (0.419)	3.653**** (0.242)	3.810**** (0.135)
Observations	2,839	697	816	2,839
Prob > χ^2	0.000	0.000	0.000	0.000
R ²	0.650	0.598	0.695	0.651

Table D2.2. Earnings as a function of individual characteristics and market variables for traders who hold accurate beliefs about their CRT rank and those who do not.

	<i>CRT beliefs market mismatch</i>	<i>CRT beliefs market mismatch</i>
Sample:	Equal to zero	Different from zero
Intercept	978.050**** (92.058)	1,071.464**** (52.046)
<i>Hypothesized predictors</i>		
CRT score	5.926 (35.507)	48.772** (19.755)
ToM score	66.926 (46.094)	34.862** (16.302)
<i>Controls</i>		
<i>Individual characteristics</i>		
Raven score	35.602 (29.780)	46.533** (21.959)
Financial literacy score	79.661 (63.647)	2.013 (12.880)
Male	99.447 (68.555)	69.137* (41.388)
<i>Market characteristics</i>		
Loan Dummy	-716.166**** (95.797)	-879.815**** (51.675)
Hint “Not 50”	162.573 (112.131)	200.650**** (51.107)
Hint “Not 490”	177.156** (79.831)	239.868**** (18.681)
Asset value	4.328**** (0.373)	3.735**** (0.147)
Observations	357	2,482
Prob > χ^2	0.000	0.000
R ²	0.719	0.642

Appendix E. Gender effects

Gender and earnings

In line with previous research, we find that men have higher CRT scores than women (3.42 vs. 2.53, p-value=0.001, WRS) (Frederick, 2005, Brañas-Garza, Kujal and Lenkei, 2016) whereas the opposite is true for ToM scores (24.41 vs. 26.61, p-value = 0.023, WRS) (Baron Cohen et al. 1997). Thus, our findings put forward that both ‘men’ and ‘women’ skills are needed to perform well in markets. However, in our setting, men tend to earn more (\$38.98) than women (\$37.21) (p-value = 0.013, WRS).

Interestingly, this male effect disappears when considering other measures of trader performance such as the proportion of consistent orders (see Tables C3.1 and C3.2) or net holdings consistency (Table C4.1). The proportion of consistent orders for men is equal to 60.72% compared to 59.67% for women (p-value = 0.690, WRS). Net holdings consistency for women (0.036) is actually slightly higher than for men (-0.054) (p-value = 0.625, WRS). In addition, in situations in which theory of mind is supposedly the most important, such as the Winner's curse traps situations à la Biais et al. (2005) (Table C1.1) or imprecise hints (Table C2.1), the Male Dummy does not explain trader earnings significantly.

Trading activity and earnings

In line with Barber and Odean (2001) and Biais et al. (2005), we find that trading activity as measured by the number of completed transactions in a given period negatively affects earnings (see Table E1).⁵⁶ Also in line with Barber and Odean (2001) and Biais et al. (2005), we find that men (6.88) trade more than women (5.87) (p-value = 0.082, WRS). Interestingly, not all men trade more than women. Instead, high trading skills (Top 25% in CRT and ToM scores) men engage in more trades than high trading skills women (7.20 vs. 4.67, p-value = 0.074, WRS) whereas the opposite is true for low trading skills (4.44 vs. 6.43, which is not significant however, p-value = 0.262, WRS). It follows that men who are in the top 25% in terms of CRT and ToM scores trade more than those who are in the bottom 25% (p-value = 0.045, WRS) whereas it is not the case for women (p-value = 0.401, WRS).

Relatedly, men are not more overconfident in their trading skills than women. According to our CRT-based overconfidence measure, men and women do not differ in terms of overconfidence (1.328 vs 1.528, p-value = 0.693, WRS). In line with Biais et al. (2005) results, we find that our measure of overconfidence does not correlate with trading activity (p-value = 0.811).

To account for the difference in trading activity between men and women, we calculate the net gains per trade for each trader in each market. This is calculated as the earnings of the trader in a given market minus the value of the portfolio each trader is endowed with (this is calculated as

⁵⁶ Similar results are obtained if we use the total number of orders as a measure of trading activity.

the cash endowment plus the number of shares each trader is given multiplied by the value of the asset in a given market) divided by the number of transactions completed by that trader. The Male Dummy coefficient becomes statistically insignificant when using net gains per trade as the dependent variable instead of earnings (see Table E1).

Table E1. Trader earnings and net gains per trade as a function of individual characteristics and market variables

	Trader earnings	Net gains per trade
Intercept	1,124.844**** (42.377)	-236.458**** (22.307)
<i>Hypothesized predictors</i>		
CRT score	46.149*** (17.262)	9.776** (4.022)
ToM score	38.801**** (11.944)	9.628** (4.071)
Number of trades	-11.467*** (3.670)	-0.642 (2.142)
<i>Controls</i>		
<i>Individual characteristics</i>		
Raven score	47.559*** (16.546)	0.862 (4.572)
Financial literacy score	12.302 (13.361)	2.681 (3.642)
Male dummy	82.398*** (30.824)	-5.161 (10.651)
<i>Market characteristics</i>		
Loan Dummy	-806.333**** (37.344)	-165.816**** (14.159)
Hint “Not 50” Dummy	204.446**** (45.884)	34.282* (19.776)
Hint “Not 490” Dummy	231.689**** (17.515)	37.125** (15.335)
Asset value	3.777**** (0.135)	1.032**** (0.091)
Observations	n = 2,839	n = 2,653
Prob > χ^2	0.000	0.000

R^2	0.658	0.442
-------	-------	-------

Appendix F. Robustness checks

Appendix F1. Additional individual controls

We add the following individual controls to our regressions: risk attitudes, GPA, school dummies, personality traits (big five and self-monitoring scale). All of these measures were collected during the one-hour follow-up survey (see Appendix A).

Risk attitudes

Controlling for risk attitudes allows us to eliminate a possible confound in our analysis of trader earnings. For example, more risk tolerant traders may be more likely to trade (Fellner and Maciejovsky, 2007) and engage in risky, speculative trading strategies that may lead to higher returns. In addition, risk aversion has been found to correlate negatively with cognitive ability (Dohmen et al. 2010; Beauchamp, Cesarini and Johannesson, 2014) so that controlling for risk attitudes appears to be essential in isolating the effect of cognitive ability on trader performance.

Our measure of an individual's risk attitudes is the number of safe choices in the Holt and Laury risk elicitation task (2002).⁵⁷

Personality traits

We administered the 44-item version of the Big Five personality test which was developed by John, Donahue and Kentle (1991) and John, Naumann and Soto (2008). We use personality traits as control variables because they have been shown to relate to financial decision making. For example, Brown and Taylor (2014) show that personality traits may affect household finance decisions identifying, for example, a positive relationship between openness and the probability of holding stocks. Also, Fernandes, Lynch and Netemeyer (2014) show that controlling for

⁵⁷ The proportion of subjects inconsistently switching between the safer and the riskier option (13.2%) was similar to the proportion reported in Holt and Laury (2002) (about 10% of the cases). Our analysis is not qualitatively affected by removing these inconsistent subjects from the analysis. In the results section we present the results for all subjects, regardless of their consistency in the Holt and Laury risk attitudes elicitation task.

personality traits affects the magnitude and significance of the effect of financial literacy on financial decisions. Finally, personality traits have been found to correlate with several of the measures collected in our study. For example, openness has been found to correlate positively with Raven test scores (see DeYoung, 2011; Beauchamp, Cesarini and Johannesson, 2012), while both agreeableness and openness have been found to correlate positively with theory of mind skills (see Mayer et al. 2011).

School dummies take value 1 for each of the following schools: business and economics (27% of the subjects), educational studies (5%), film and media arts (19%), humanities and social sciences (14%), law school (8%), performing arts (8%), science and technology (15%), pharmacy and health and behavioral sciences (4%).

The self-monitoring scale is based on Snyder and Gangestad (1986) which is the scale used by Biais et al. (2005).

Table F1.1. Earnings as a function of extended individual characteristics and market variables

Sample:	All	Top 25% CRT	Bottom 25% CRT
Intercept	1,068.356**** (182.569)	1.782 (244.548)	1,485.138**** (316.070)
<i>Hypothesized predictors</i>			
CRT score	47.878*** (17.381)	81.938 (53.185)	39.937 (91.657)
ToM score	33.568**** (9.966)	107.153**** (21.323)	28.356 (22.769)
<i>Controls</i>			
<i>Individual characteristics</i>			
Raven score	41.615** (17.491)	40.110** (18.035)	43.273** (21.918)
Financial literacy score	-7.653 (15.394)	-35.758** (17.816)	28.829 (26.594)
Male	61.233* (34.740)	99.832** (45.934)	126.561 (87.653)
Risk attitudes	-1.009 (7.554)	2.204 (10.414)	-0.049 (9.277)
GPA	30.481 (30.713)	338.003**** (75.714)	59.581 (55.376)

School Dummies	ns	ns ⁵⁸	ns ⁵⁹
Big five personality traits & Self-monitoring scale	ns	ns	ns
<i>Market characteristics</i>			
Loan Dummy	-846.254**** (40.970)	-922.309**** (71.175)	-710.814**** (78.852)
Hint “Not 50”	197.531**** (45.185)	131.290 (79.843)	222.934**** (63.918)
Hint “Not 490”	233.167**** (16.454)	175.287**** (33.607)	215.877**** (46.986)
Asset value	3.808**** (0.138)	4.014**** (0.180)	3.761**** (0.252)
Observations	2,839	1,020	816
Prob > χ^2	0.000	0.000	0.000
R ²	0.655	0.687	0.679

Table F1.2. Proportion of consistent orders and net holdings consistency as a function of extended individual characteristics and market variables

Sample:	Proportion of consistent orders		Net holdings consistency	
	All	Top 25% CRT	All	Top 25% CRT
Intercept	0.351*** (0.111)	0.003 (0.183)	0.000 (0.992)	-8.244**** (1.195)
<i>Hypothesized predictors</i>				
CRT score	0.038**** (0.008)	0.070*** (0.023)	0.350*** (0.123)	0.764*** (0.287)
ToM score	0.011 (0.010)	0.053**** (0.014)	0.079 (0.080)	0.711**** (0.136)
<i>Controls</i>				
<i>Individual characteristics</i>				
Raven score	0.016** (0.008)	0.023* (0.012)	0.081 (0.086)	0.379**** (0.115)
Financial literacy score	0.013 (0.008)	0.001 (0.012)	0.117 (0.100)	0.002 (0.117)
Male	0.002	0.037	0.024	0.795***

⁵⁸ We observe a significant (p-value = 0.043) positive effect for the dummy variable associated to the school of Business and Economics.

⁵⁹ We observe a significant (p-value = 0.002) positive effect for the dummy variable associated to the school of Business and Economics.

	(0.014)	(0.038)	(0.231)	(0.262)
Risk attitudes	-0.002 (0.005)	-0.005 (0.007)	-0.003 (0.034)	0.093 (0.076)
GPA	0.007 (0.011)	0.104*** (0.038)	-0.134 (0.172)	2.292*** (0.372)
School Dummies	ns	ns	ns	ns ⁶⁰
Big five personality traits & Self-monitoring scale	ns	ns	ns	ns
<i>Market characteristics</i>				
Loan Dummy	0.021 (0.018)	0.001 (0.050)	0.447*** (0.167)	0.816*** (0.290)
High stakes Dummy	-0.017 (0.024)	0.007 (0.036)	0.314 (0.198)	0.300 (0.477)
Hint “Not 50”	0.148*** (0.029)	0.165*** (0.040)	1.904*** (0.448)	1.893** (0.900)
Hint “Not 490”	0.180*** (0.028)	0.182*** (0.037)	1.928*** (0.477)	2.178*** (0.812)
Asset value	-0.001** (0.001)	-0.001 (0.002)	-0.001 (0.001)	0.002 (0.002)
Consistent with hint	0.197*** (0.021)	0.171*** (0.047)	—	—
Number of transactions	-0.006*** (0.001)	-0.006*** (0.001)	-0.050* (0.026)	-0.039*** (0.014)
Market number	-0.002**	-0.001	—	—
Observations	2,698	1,000	1,670	600
Prob > χ^2	0.000	0.000	0.000	0.000
R ²	0.167	0.182	0.0685	0.100

Appendix F2. Session fixed effects

Tables F2.1 and F2.2 report results for regressions that include session fixed effects.

Table F2.1. Earnings as a function of individual characteristics and market variables

Sample:	All	Top 25% CRT	Bottom 25% CRT
Intercept	1,103.892***	898.194***	1,105.777***

⁶⁰ We observe a significant (p-values < 0.05) positive effect for the dummy variables associated to the school of Humanities and for the school of Health and Behavioral Sciences.

	(42.702)	(98.288)	(185.452)
<i>Hypothesized predictors</i>			
CRT score	50.956*** (18.319)	181.484** (73.069)	9.525 (148.442)
ToM score	36.203*** (12.801)	61.650** (26.862)	57.909** (24.388)
<i>Controls</i>			
<i>Individual characteristics</i>			
Raven score	45.523** (18.233)	36.422 (32.185)	43.153* (25.249)
Financial literacy score	3.153 (18.705)	11.510 (28.353)	18.924 (31.031)
Male	84.366** (38.471)	27.195 (78.884)	155.758** (74.216)
<i>Market characteristics</i>			
Loan Dummy	-881.650***** (25.505)	-928.579***** (40.818)	-691.640***** (36.115)
Hint “Not 50”	198.821***** (45.905)	139.237* (76.284)	223.509***** (60.944)
Hint “Not 490”	234.017***** (16.375)	167.845***** (35.215)	217.207***** (47.025)
Asset value	3.807***** (0.136)	3.997***** (0.181)	3.762***** (0.248)
Observations	2,839	1,020	816
R ²	0.652	0.679	0.675

Table F2.2. Proportion of consistent orders and net holdings consistency as a function of individual characteristics and market variables

Sample:	Proportion of consistent orders		Net holdings consistency	
	All	Top 25% CRT	All	Top 25% CRT
Intercept	0.465***** (0.031)	0.399***** (0.054)	-0.547** (0.259)	-1.921* (1.152)
<i>Hypothesized predictors</i>				
CRT score	0.032***** (0.008)	0.094***** (0.026)	0.301*** (0.096)	1.274** (0.563)
ToM score	0.006 (0.009)	0.044*** (0.015)	0.007 (0.060)	0.396* (0.239)
<i>Controls</i>				
<i>Individual characteristics</i>				
Raven score	0.015** (0.007)	0.012 (0.017)	0.095 (0.090)	0.351 (0.300)

Financial literacy score	0.015** (0.007)	0.004 (0.015)	0.115 (0.125)	0.061 (0.219)
Male	-0.003 (0.010)	-0.001 (0.038)	-0.222 (0.239)	-0.257 (0.417)
<i>Market characteristics</i>				
Loan Dummy	-0.008 (0.009)	-0.083**** (0.023)	-0.072 (0.138)	-0.931** (0.384)
High stakes Dummy	-0.059**** (0.008)	-0.110**** (0.017)	0.221 (0.151)	-0.618 (0.389)
Hint “Not 50”	0.148**** (0.028)	0.167**** (0.038)	1.943**** (0.493)	1.903* (1.041)
Hint “Not 490”	0.180**** (0.028)	0.175**** (0.036)	1.888**** (0.445)	2.044**** (0.678)
Asset value	-0.001** (0.0001)	-0.001 (0.001)	-0.001 (0.001)	0.002 (0.002)
Consistent with hint	0.197**** (0.022)	0.174**** (0.050)	—	—
Number of transactions	-0.006**** (0.001)	-0.006**** (0.001)	-0.051 (0.031)	-0.060**** (0.018)
Market number	-0.002* (0.001)	-0.001 (0.002)	—	—
Observations	2,698	1,000	1,670	600
R ²	0.164	0.170	0.0600	0.0696

Appendix F3. Analysis of total earnings and total proportion of consistent orders

We define a subject’s total earnings as the sum of her earnings from all seventeen markets. Total consistency of orders is defined as the average proportion of consistent orders of a given trader over the course of the seventeen markets.

Table F3.1. Earnings and proportion of consistent orders as a function of individual characteristics and market variables

Sample:	Earnings			Proportion of consistent orders	
	All	Top 25% CRT	Bottom 25% CRT	All	Top 25% CRT
Intercept	37,216.955**** (269.417)	35,541.383**** (1,134.972)	35,657.499**** (2,322.039)	0.475**** (0.106)	0.567* (0.269)
<i>Hypothesized predictors</i>					
CRT score	707.776** (306.160)	2,546.296** (1,122.593)	-106.951 (1,974.757)	0.019 (0.028)	-0.031 (0.066)
ToM score	658.585**	1,339.453***	821.068**	0.026	0.076**

	(261.239)	(425.966)	(351.592)	(0.023)	(0.032)
<i>Controls</i>					
<i>Individual characteristics</i>					
Raven score	773.014** (336.221)	704.426 (518.552)	875.471* (437.662)	0.030* (0.016)	0.026 (0.039)
Financial literacy score	150.507 (255.266)	-113.119 (409.257)	666.528 (434.013)	0.045* (0.022)	0.051 (0.037)
Male	1,227.010* (643.393)	296.111 (1,027.378)	2,571.782** (1,015.524)	0.066 (0.048)	0.163* (0.086)
<i>Market characteristics</i>					
Loan Dummy	-14,702.507**** (545.940)	-16,603.597**** (738.950)	-11,441.226**** (1,228.827)	0.062** (0.028)	-0.112* (0.056)
High stakes Dummy	—	—	—	-0.012 (0.036)	-0.114 (0.106)
Consistent with hint	—	—	—	0.408*** (0.130)	0.375 (0.291)
Number of transactions	—	—	—	-0.014**** (0.003)	-0.014* (0.007)
Observations	167	60	48	155	57
R ²	0.689	0.743	0.621	0.287	0.345

Even though panel data databases with less than 20 time observations (17 in our case) are less subject to serial correlation issues (Baltagi, 2013), we assessed the robustness of our findings to allowing for an autoregressive structure of order 1 in the error term (see Tables F3.2 and F3.3). Similar results are obtained if we allow for a moving average of order 1 instead.

Table F3.2. Earnings as a function of individual characteristics and market variables

	All	Top 25% CRT	Bottom 25% CRT
Intercept	1,063.802**** (29.817)	964.354**** (66.292)	988.254**** (137.739)
<i>Hypothesized predictors</i>			
CRT score	45.047*** (15.473)	143.872*** (48.135)	7.159 (105.925)
ToM score	39.072** (15.278)	77.781** (32.596)	46.035** (23.005)
<i>Controls</i>			
<i>Individual characteristics</i>			
Raven score	45.241*** (15.028)	42.240 (29.110)	48.525** (22.278)
Financial literacy score	8.588	-4.902	35.167

	(15.036)	(25.635)	(27.229)
Male	75.574**	24.864	150.441**
	(32.864)	(54.477)	(61.599)
<i>Individual characteristics</i>			
Loan Dummy	-869.292****	-974.800****	-673.388****
	(46.949)	(77.043)	(94.883)
Hint “Not 50”	197.337****	134.671***	220.924****
	(24.573)	(41.394)	(44.990)
Hint “Not 490”	224.863****	161.873****	211.717****
	(24.398)	(41.483)	(44.028)
Asset value	3.786****	3.984****	3.742****
	(0.062)	(0.105)	(0.115)
Observations	2,839	1,020	816
Prob > χ^2	0.000	0.000	0.000
R ²	0.650	0.670	0.661

Table F3.3. Proportion of consistent orders and net holdings consistency as a function of individual characteristics and market variables

Sample:	Proportion of consistent orders		Net holdings consistency	
	All	Top 25% CRT	All	Top 25% CRT
Intercept	0.439**** (0.023)	0.379**** (0.043)	-0.875** (0.340)	-2.551*** (0.794)
<i>Hypothesized predictors</i>				
CRT score	0.033**** (0.007)	0.082**** (0.023)	0.284* (0.150)	1.173** (0.537)
ToM score	0.010 (0.007)	0.043*** (0.016)	0.016 (0.146)	0.526 (0.363)
<i>Controls</i>				
<i>Individual characteristics</i>				
Raven score	0.016** (0.007)	0.012 (0.014)	0.093 (0.143)	0.391 (0.324)
Financial literacy	0.014** (0.007)	0.009 (0.012)	0.072 (0.144)	0.022 (0.285)
Male	-0.003 (0.015)	-0.003 (0.027)	-0.272 (0.315)	-0.195 (0.615)
<i>Market characteristics</i>				
Loan Dummy	0.012 (0.022)	-0.044 (0.040)	0.122 (0.474)	0.168 (0.922)

High stakes Dummy	-0.016 (0.015)	-0.022 (0.027)	0.221 (0.315)	0.180 (0.625)
Hint “Not 50”	0.151**** (0.015)	0.166**** (0.025)	2.046**** (0.350)	2.012*** (0.701)
Hint “Not 490”	0.181**** (0.015)	0.177**** (0.026)	1.898**** (0.318)	2.067*** (0.646)
Asset value	-0.001**** (0.0001)	-0.001* (0.001)	-0.001 (0.002)	0.002 (0.001)
Consistent with hint	0.193**** (0.020)	0.179**** (0.035)	—	—
Number of transactions	-0.006**** (0.001)	-0.006**** (0.002)	-0.024 (0.020)	-0.042 (0.036)
Observations	2,698	1,000	1,670	600
Prob > χ^2	0.000	0.000	0.000	0.001
R ²	0.161	0.157	0.0565	0.0651

Appendix F4. Non-orthogonal variables

Table F4.1. Earnings as a function of individual characteristics and market variables

Sample:	All	All	Top 25% CRT	Bottom 25% CRT
Intercept	809.640**** (158.532)	605.696**** (138.689)	221.846 (304.929)	519.199**** (190.366)
<i>Hypothesized predictors</i>				
CRT score	17.081** (11.596)	15.030 (11.596)	71.821** (33.340)	-14.526 (60.721)
ToM score	—	4.672** (2.103)	11.531**** (3.531)	5.136* (2.864)
<i>Controls</i>				
<i>Individual characteristics</i>				
Raven score	15.065 (9.657)	19.399** (8.087)	18.228 (13.084)	19.936* (10.261)
Financial literacy score	1.797 (5.095)	3.480 (6.014)	-2.121 (9.571)	14.751 (10.806)
Male	64.802* (37.665)	74.916** (35.923)	24.505 (56.777)	147.881**** (52.012)
<i>Market characteristics</i>				
Loan Dummy	-879.488**** (23.809)	-865.255**** (32.963)	-971.785**** (42.289)	-663.644**** (73.364)
Hint “Not 50”	191.144**** (48.822)	198.245**** (45.523)	137.586* (75.778)	222.133**** (59.192)
Hint “Not 490”	211.048****	234.134****	169.026****	218.423****

	(20.222)	(16.264)	(34.530)	(46.904)
Asset value	3.835****	3.808****	4.000****	3.765****
	(0.132)	(0.135)	(0.180)	(0.243)
Observations	3,468	2,839	1,020	816
Prob > χ^2	0.000	0.000	0.000	0.000
R ²	0.645	0.650	0.670	0.661

Table F4.2. Proportion of consistent orders and net holdings consistency as a function of individual characteristics and market variables

Sample:	Proportion of consistent orders		Net holdings consistency	
	All	Top 25% CRT	All	Top 25% CRT
Intercept	0.264**** (0.047)	0.008 (0.133)	-1.873*** (0.659)	-8.175** (3.399)
<i>Hypothesized predictors</i>				
CRT score	0.012** (0.005)	0.041*** (0.013)	0.104* (0.060)	0.529** (0.227)
ToM score	0.001 (0.002)	0.007** (0.003)	0.002 (0.011)	0.075*** (0.026)
<i>Controls</i>				
<i>Individual characteristics</i>				
Raven score	0.007** (0.003)	0.005 (0.006)	0.032 (0.037)	0.156 (0.113)
Financial literacy score	0.006** (0.003)	0.004 (0.005)	0.045 (0.043)	0.021 (0.066)
Male	-0.002 (0.010)	-0.003 (0.027)	-0.149 (0.187)	-0.096 (0.266)
<i>Market characteristics</i>				
Loan Dummy	0.014 (0.010)	-0.043** (0.019)	0.317** (0.159)	0.271 (0.366)
High stakes Dummy	-0.016 (0.010)	-0.022 (0.026)	0.241*** (0.085)	0.147 (0.185)
Hint “Not 50”	0.148**** (0.028)	0.167**** (0.037)	1.920**** (0.473)	1.931** (0.889)
Hint “Not 490”	0.180**** (0.028)	0.175**** (0.036)	1.901**** (0.450)	2.049**** (0.696)
Asset value	-0.001** (0.0001)	-0.001 (0.001)	-0.001 (0.001)	0.002 (0.002)
Consistent with hint	0.197**** (0.022)	0.176**** (0.051)	—	—
Number of transactions	-0.006**** (0.001)	-0.006**** (0.001)	-0.047* (0.028)	-0.050**** (0.014)

Market number	-0.002** (0.001)	-0.001 (0.002)	—	—
Observations	2,698	1,000	1,670	600
Prob > χ^2	0.000	0.000	0.000	0.000
R ²	0.161	0.157	0.0580	0.0654

Appendix F5. High stakes sessions

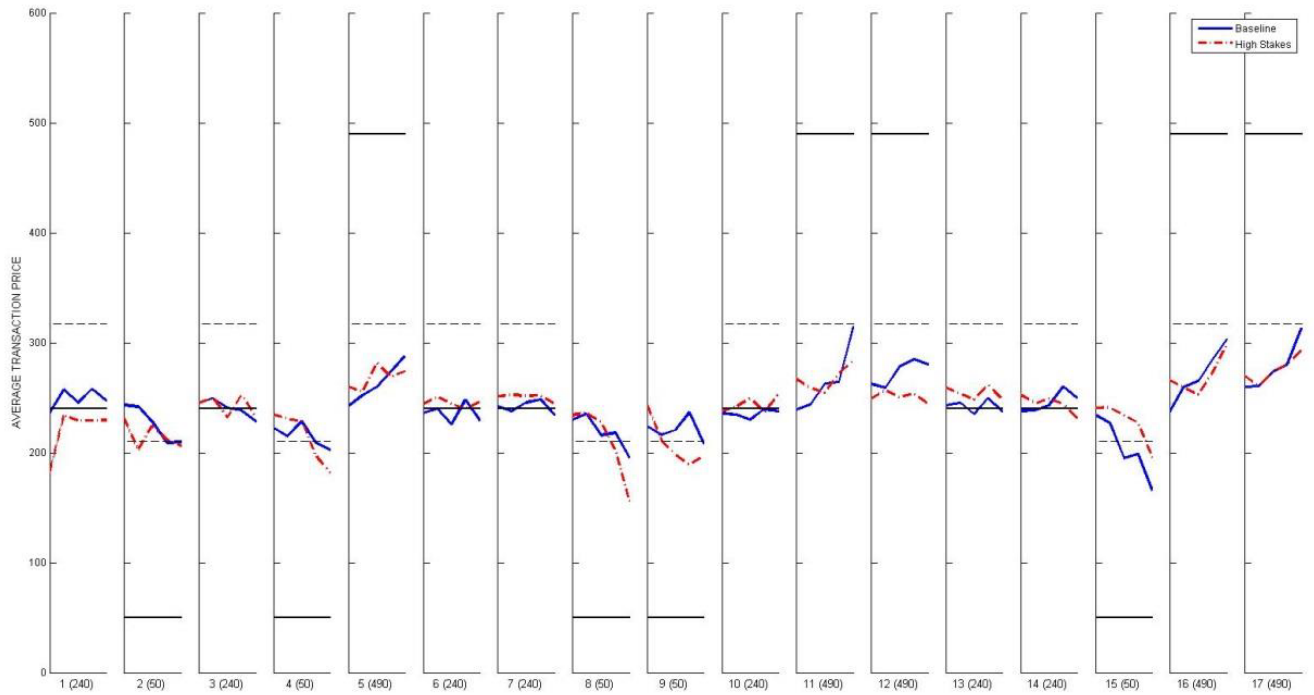


Figure F5.1. Average prices per minute per market for the ten baseline and five high stakes sessions (dashed curves). Prices for the high stakes sessions are divided by two to facilitate comparison between the two treatments.

Table F5.1. Earnings as a function of individual characteristics and market variables

Sample:	All	Top 25%	Bottom 25%
---------	-----	---------	------------

		CRT	CRT
Intercept	1,030.443**** (48.222)	963.502**** (102.196)	797.605*** (269.070)
<i>Hypothesized predictors</i>			
CRT score	47.323*** (18.143)	130.732 (86.296)	-75.113 (18.143)
ToM score	40.604** (15.867)	71.252*** (23.842)	59.895 (39.656)
CRT score × High stakes Dummy	6.605 (40.138)	51.101 (104.431)	131.627 (239.143)
ToM score × High stakes Dummy	-11.348 (22.661)	14.833 (90.536)	-42.465 (46.962)
<i>Controls</i>			
<i>Individual characteristics</i>			
Raven score	45.051** (17.722)	43.212 (29.865)	42.174* (22.222)
Financial literacy score	6.609 (15.582)	-2.213 (23.152)	34.125 (20.527)
Male Dummy	78.508** (36.632)	24.063 (59.144)	-572.762**** (73.948)
<i>Market characteristics</i>			
Loan Dummy	-843.291**** (32.719)	-959.753**** (43.284)	-575.762**** (73.948)
High stakes Dummy	58.096**** (13.999)	-26.298 (102.044)	289.274 (281.067)
Hint “Not 50” Dummy	198.673**** (45.554)	138.112**** (75.715)	224.101**** (58.775)
Hint “Not 490” Dummy	234.333**** (13.313)	169.075**** (34.601)	2184.926**** (47.440)
Asset value	3.808**** (0.136)	4.000**** (0.180)	3.763**** (0.243)
Observations	2,839	1,020	816
Prob > χ^2	0.000	0.000	0.000
R ²	0.651	0.670	0.668

Table F5.2. Proportion of consistent orders and net holdings consistency as a function of individual characteristics and market variables

Proportion of consistent orders	Net holdings consistency
---------------------------------	--------------------------

Sample:	All	Top 25% CRT	All	Top 25% CRT
Intercept	0.450**** (0.031)	0.392**** (0.068)	-0.785*** (0.277)	-2.331* (1.336)
<i>Hypothesized predictors</i>				
CRT score	0.028*** (0.008)	0.078** (0.033)	0.225* (0.127)	0.969 (0.606)
ToM score	0.015 (0.011)	0.047**** (0.014)	0.038 (0.081)	0.499** (0.234)
CRT score \times High stakes Dummy	0.0156 (0.017)	0.040 (0.038)	0.113 (0.162)	0.519 (0.710)
ToM score \times High stakes Dummy	-0.022 (0.015)	0.081* (0.043)	-0.037 (0.142)	-0.057 (0.578)
<i>Controls</i>				
<i>Individual characteristics</i>				
Raven score	0.014** (0.007)	0.009 (0.013)	0.076 (0.086)	0.365 (0.246)
Financial literacy score	0.014** (0.007)	0.012 (0.013)	0.112 (0.106)	0.071 (0.171)
Male Dummy	-0.001 (0.011)	-0.003 (0.028)	-0.152 (0.194)	-0.063 (0.284)
<i>Market characteristics</i>				
Loan Dummy	0.017* (0.010)	0.041** (0.018)	0.324** (0.163)	0.320 (0.350)
High stakes Dummy	-0.011 (0.010)	-0.044 (0.043)	0.260*** (0.088)	-0.358 (0.761)
Hint “Not 50” Dummy	0.148**** (0.028)	0.168**** (0.037)	1.927**** (0.471)	1.941** (0.869)
Hint “Not 490” Dummy	0.180**** (0.028)	0.176**** (0.036)	1.899**** (0.452)	2.053*** (0.695)
Asset value	-0.001** (0.001)	-0.001** (0.001)	-0.001 (0.002)	-0.001 (0.002)
Consistent with hint	0.197**** (0.022)	0.175*** (0.051)	—	—
Number of transactions	-0.006**** (0.001)	-0.006**** (0.001)	—	—
Market number	-0.002** (0.001)	-0.001 (0.002)	—	—
Observations	2,698	1,000	1,670	600
Prob > χ^2	0.000	0.000	0.000	0.000
R ²	0.162	0.160	0.058	0.066

Appendix F6. Independent variables measured as deviations from session average⁶¹

Table F6.1. Earnings as a function of individual characteristics and market variables

Sample:	All	Top 25% CRT	Bottom 25% CRT
Intercept	1,055.636**** (45.909)	1,011.268**** (77.775)	1,119.415**** (81.394)
<i>Hypothesized predictors</i>			
CRT score	50.784*** (18.456)	38.645 (25.801)	62.433*** (22.289)
ToM score	49.484*** (16.871)	106.303** (53.006)	135.296** (57.107)
<i>Controls</i>			
<i>Individual characteristics</i>			
Raven score	8.312 (17.465)	18.863 (24.718)	29.777 (26.087)
Financial literacy score	23.323** (11.657)	29.333 (25.191)	34.823* (18.597)
Male	69.553** (34.579)	17.490 (54.342)	143.237*** (49.951)
<i>Market characteristics</i>			
Loan Dummy	-859.578**** (28.433)	-964.161**** (41.366)	-636.480**** (86.397)
Hint “Not 50”	198.967**** (45.674)	140.310* (76.254)	220.197**** (57.752)
Hint “Not 490”	234.146**** (16.321)	168.796**** (34.662)	219.408**** (47.193)
Asset value	3.807**** (0.136)	3.997**** (0.180)	3.769**** (0.241)
Observations	2,839	1,020	816
Prob > χ^2	0.000	0.000	0.000
R ²	0.650	0.667	0.665

Table F6.2. Proportion of consistent orders and net holdings consistency as a function of individual characteristics and market variables

Sample:	Proportion of consistent orders		Net holdings consistency	
	All	Top 25% CRT	All	Top 25% CRT
Intercept	0.457**** (0.030)	0.424**** (0.068)	-0.735*** (0.256)	-2.075** (1.027)
<i>Hypothesized predictors</i>				

⁶¹ Similar results are obtained when using session median instead.

CRT score	0.015*** (0.005)	0.009 (0.010)	0.091 (0.088)	0.360 (0.244)
ToM score	0.032*** (0.008)	0.065*** (0.022)	0.273*** (0.086)	0.969*** (0.354)
<i>Controls</i>				
<i>Individual characteristics</i>				
Raven score	0.016** (0.007)	0.019 (0.013)	0.099 (0.108)	0.164 (0.190)
Financial literacy score	0.002 (0.009)	0.021 (0.020)	-0.008 (0.057)	0.257* (0.154)
Male	-0.005 (0.010)	-0.004 (0.025)	-0.169 (0.189)	-0.125 (0.236)
<i>Market characteristics</i>				
Loan Dummy	0.013 (0.008)	-0.053*** (0.015)	0.319** (0.152)	0.202 (0.321)
High stakes Dummy	-0.031** (0.013)	-0.048 (0.029)	0.126 (0.081)	-0.268 (0.259)
Hint “Not 50”	0.148*** (0.028)	0.170*** (0.037)	1.930*** (0.473)	1.966** (0.914)
Hint “Not 490”	0.179*** (0.028)	0.175*** (0.036)	1.897*** (0.451)	2.033*** (0.689)
Asset value	-0.001** (0.0001)	-0.001 (0.001)	-0.001 (0.001)	0.002 (0.002)
Consistent with hint	0.198*** (0.021)	0.174*** (0.049)	—	—
Number of transactions	-0.006*** (0.001)	-0.006*** (0.001)	-0.048* (0.028)	-0.054*** (0.015)
Market number	-0.002** (0.001)	-0.001 (0.002)	—	—
Observations	2,698	1,000	1,670	600
Prob > χ^2	0.000	0.000	0.000	0.000
R ²	0.160	0.151	0.0583	0.0638

Online Appendix

Appendix O1 (instructions) and O2 (asset prices charts). (click [here](#))

Link: goo.gl/PBLxWs